



**TUGAS AKHIR - SS141501**

**DETEKSI ANOMALI PADA PEMAKAIAN AIR  
PELANGGAN PDAM SURYA SEMBADA KOTA  
SURABAYA MENGGUNAKAN ALGORITMA *KOHONEN  
SELF ORGANIZING MAPS (SOM)* DAN *LOCAL OUTLIER  
FACTOR (LOF)***

**AMRI MUHAIMIN  
NRP 062114 4000 0125**

**Dosen Pembimbing  
Dr. Kartika Fithriasari, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**



**TUGAS AKHIR - SS141501**

**DETEKSI ANOMALI PADA PEMAKAIAN AIR  
PELANGGAN PDAM SURYA SEMBADA KOTA  
SURABAYA MENGGUNAKAN ALGORITMA *KOHONEN  
SELF ORGANIZING MAPS (SOM)* DAN *LOCAL  
OUTLIER FACTOR (LOF)***

**AMRI MUHAIMIN  
NRP 062114 4000 0125**

**Dosen Pembimbing  
Dr. Kartika Fithriasari, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**





**FINAL PROJECT - SS 141501**

***ANOMALY DETECTION ON CUSTOMER WATER  
CONSUMPTION AT LOCAL WATER DRINKING  
COMPANY OF SURABAYA CITY USING  
KOHONEN SELF ORGANIZING MAPS (SOM) AND  
LOCAL OUTLIER FACTOR (LOF)***

**AMRI MUHAIMIN  
SN 062114 4000 0125**

**Supervisor  
Dr. Kartika Fithriasari, M.Si**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2018**





## LEMBAR PENGESAHAN

# DETEKSI ANOMALI PADA PEMAKAIAN AIR PELANGGAN PDAM SURYA SEMBADA KOTA SURABAYA MENGGUNAKAN ALGORITMA *KOHONEN SELF ORGANIZING MAPS (SOM) DAN LOCAL OUTLIER FACTOR (LOF)*

### TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Sains  
pada  
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

**Amri Muhaimin**

NRP. 062114 4000 0125

Disetujui oleh Pembimbing:  
**Dr. Kartika Fithriasari**  
NIP. 19691212 199303 2 002

(  )



Mengetahui,  
Kepala Departemen

**Dr. Suhartono**

NIP. 19710929199512 1 0014

SURABAYA, JULI 2018





**DETEKSI ANOMALI PADA PEMAKAIAN AIR  
PELANGGAN PDAM SURYA SEMBADA KOTA  
SURABAYA MENGGUNAKAN ALGORITMA KOHONEN  
SELF-ORGANIZING MAPS (SOM) DAN LOCAL OUTLIER  
FACTOR (LOF)**

**Nama Mahasiswa : Amri Muhaimin**  
**NRP : 06211440000125**  
**Departemen : Statistika**  
**Dosen Pembimbing : Dr. Kartika Fithriasari, M.Si**

**Abstrak**

*Kehilangan air dalam distribusi merupakan masalah yang cukup serius, kerugian yang disebabkan oleh tingkat kehilangan PDAM Surya Sembada Kota Surabaya mencapai 2 Miliar Rupiah. Pada tahun 2011 terjadi pencurian air sebanyak 100 kasus. Deteksi anomali pada penelitian dilakukan menggunakan data selama Maret 2017 – Februari 2018. Data yang digunakan adalah pemakaian air kemudian didapatkan variabel-variabel rata-rata pemakaian air, maksimal pemakaian air, dan deviasi standar pemakaian air. Algoritma Kohonen-SOM mendapatkan 45 kelompok yang dianggap kelompok anomali dengan kriteria silhouette width kurang dari rata-rata silhouette width pada kelompok yang ter-bentuk. Terdapat 45 kelompok yang terduga anomali. Local Outlier Factor menghasilkan 1229 kejadian konsumsi yang tidak normal, 1229 kejadian tersebut terdiri dari 579 rumah tangga atau pelanggan. Perhitungan frekuensi yang dilakukan mendapatkan 42 pelanggan yang terduga anomali. Hasil deteksi anomali dengan metode PDAM dari 42 pelanggan tersebut hanya 16 yang terdeteksi. Hal tersebut dikarenakan metode PDAM gagal menangkap perilaku konsumsi yang aneh seperti konsumsi yang konstan setiap bulan. Karakteristik pelanggan yang terdeteksi anomali adalah mempunyai rata-rata pemakaian lebih dari rata-rata pemakaian golongan dan sub-zona.*

**Kata Kunci : Anomali, Kohonen-SOM, Local Outlier Factor**

*(Halaman sengaja dikosongkan)*

# **ANOMALY DETECTION ON CUSTOMER WATER CONSUMPTION AT LOCAL WATER DRINKING COMPANY OF SURABAYA CITY USING KOHONEN SELF ORGANIZING MAPS (SOM) AND LOCAL OUTLIER FACTOR (LOF)**

**Name** : Amri Muhaimin  
**Student Number** : 06211440000125  
**Department** : Statistics  
**Supervisor** : Dr. Kartika Fithriasari, M.Si

## **Abstract**

*The loss of water in the distribution is a serious problem in PDAM Surabaya, the loss caused by loss rate of PDAM Surabaya reach 2 billion Rupiah. In 2011 there was a theft of 100 cases. Detection of anomalies in the study was conducted using data during March 2017 - February 2018. The data used is water consumption then obtained the variables like mean of water use, maximum water use, and standart deviation of water usage. The Kohonen-SOM algorithm obtained 45 groups considered anomalous group with silhouette width criteria less than the silhouette width average in the group formed. There are 45 groups of unexpected anomalies. Local Outlier Factor produced 1229 unusual consumption events, 1229 incidents consisting of 579 households or customers. The frequency calculation performed gets 42 suspected anomaly customers. The result of anomaly detection with PDAM method from 42 customers was only 16 detected. This is because the PDAM method fails to capture strange consumption behaviors such as cons-tant consumption every month. The characteristic of the customer detected by the anomaly is to have an average of more than average usage of classes and sub-zones.*

**Keywords:** *Anomaly, Kohonen-SOM, Local Outlier Factor*

*(Halaman sengaja dikosongkan)*

## KATA PENGANTAR

Puji syukur penulis panjatkan atas rahmat dan hidayah yang diberikan Allah SWT sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Deteksi Anomali Pada Pemakaian Air Pelanggan PDAM Surya Sembada Kota Surabaya Menggunakan Kohonen SOM dan Local Outlier Factor”** dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada :

1. Pihak PDAM Surya Sembada Kota Surabaya, Pak Wawan, Pak Rian, dan Pak Nanang yang telah membimbing dan mengarahkan dalam mendapatkan data.
2. Ibu Dr. Kartika Fithriasari, M.Si selaku dosen pembimbing Tugas Akhir, yang telah meluangkan waktu dan dengan sangat sabar memberikan bimbingan, saran, dukungan serta motivasi selama penyusunan Tugas Akhir.
3. Ayah dan ibu, atas segala doa, nasehat, kasih sayang, dan dukungan yang diberikan kepada penulis demi kesuksesan dan kebahagiaan penulis, serta adikku yang selalu menghibur penulis.
4. Ibu Irhamah, M.Si, P.hD. dan Ibu Pratnya Paramitha O, M.Sc selaku dosen penguji yang telah banyak memberi masukan kepada penulis.
5. Dr. Suhartono selaku Kepala Departemen Statistika dan Dr. Sutikno, M.Si selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas, sarana, dan prasarana.
6. Mbak Saidah yang selalu memberi saya saran dan masukan.
7. Semua sahabat saya yang tidak dapat disebutkan satu per satu.

Besar harapan penulis untuk mendapatkan kritik dan saran yang membangun sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang terkait.

Surabaya, Juli 2018

Penulis

*(Halaman sengaja dikosongkan)*

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	Error! Bookmark not defined.
<b>COVER PAGE</b> .....	<b>iv</b>
<b>Abstrak</b> .....	<b>vii</b>
<b>Abstract</b> .....	<b>ix</b>
<b>KATA PENGANTAR</b> .....	<b>xi</b>
<b>DAFTAR ISI</b> .....	<b>xiii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xv</b>
<b>DAFTAR TABEL</b> .....	<b>xix</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xxi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>7</b>
2.1 Deteksi Anomali .....	7
2.2 Kohonen Self-Organizing Maps (SOM) .....	9
2.3 Local Outlier Factor.....	12
2.4 Koefisien <i>Silhouette</i> .....	14
2.5 <i>Pseudo-F Statistics</i> .....	15
2.6 Anomali Pada Konsumsi Air .....	17
2.7 Penelitian Sebelumnya .....	17
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>19</b>
3.1 Sumber Data .....	19
3.2 Variabel Penelitian .....	19
3.3 Struktur Data .....	20
3.4 Langkah Analisis .....	21
<b>BAB IV ANALISIS DAN PEMABAHASAN</b> .....	<b>25</b>
4.1 <i>Pre-Processing</i> .....	25
4.2 Karakteristik Data Pelanggan PDAM Surya Sembada Kota Surabaya.....	28
4.3 Pengelompokkan Menggunakan <i>Kohonen-SOM</i> .....	31

4.4	Pemilihan Kelompok Anomali .....	33
4.5	Deteksi Anomali Menggunakan <i>Local Outlier Factor</i> .....	69
4.6	Hasil Deteksi Anomali Dengan Metode PDAM.....	73
4.7	Karakteristik Pelanggan Yang Terdeteksi Anomali.....	74
<b>BAB V KESIMPULAN DAN SARAN.....</b>		<b>79</b>
5.1	Kesimpulan.....	79
5.2	Saran.....	79
<b>DAFTAR PUSTAKA .....</b>		<b>81</b>
<b>LAMPIRAN.....</b>		<b>85</b>
<b>BIODATA PENULIS .....</b>		<b>113</b>



## DAFTAR GAMBAR

<b>Gambar 2.1</b>	Perbedaan Anomali dan <i>Noise</i> .....	8
<b>Gambar 2.2</b>	Jenis Topologi <i>Gridtop</i> dan <i>Hextop</i> .....	10
<b>Gambar 2.3</b>	Ilustrasi Outlier .....	13
<b>Gambar 3.1</b>	Ilustrasi Nilai <i>Silhouette Width</i> Terhadap Sebaran Data.....	22
<b>Gambar 3.2</b>	Diagram Alir Penelitian .....	24
<b>Gambar 4.1</b>	Jumlah Pelanggan PDAM Surya Sembada Kota Surabaya Periode Maret 2017-Februari 2108 .....	26
<b>Gambar 4.2</b>	Persentase Pelanggan PDAM Surya Sembada Kota Surabaya Berdasarkan Zona Periode April 2017 – Februari 2018 .....	28
<b>Gambar 4.3</b>	Rata-rata Pemakaian Pelanggan Berdasarkan Zona Periode April 2017 hingga Februari 2018.....	29
<b>Gambar 4.4</b>	Rata-rata Pemakaian Air PDAM Surya Sembada Kota Surabaya Berdasarkan Golongan Tarif. ....	30
<b>Gambar 4.5</b>	Rata-rata Pemakaian Air PDAM Surya Sembada Kota Surabaya Periode Maret 2017 – Februari 2018.....	31
<b>Gambar 4.6</b>	Nilai <i>Silhouette Width</i> 9 Klaster Periode Mar 17 – Feb 18 .....	34
<b>Gambar 4.7</b>	Rata-rata Kelompok Periode Maret 2017 – Februari 2018.....	35
<b>Gambar 4.8</b>	Deviasi Standar Kelompok Periode Maret 2017 – Februari 2018 .....	36
<b>Gambar 4.9</b>	Nilai <i>Silhouette Width</i> 10 Klaster Periode Apr 17 – Feb 18.....	37
<b>Gambar 4.10</b>	Rata-rata Kelompok Periode April 2017 – Februari 2018.....	38
<b>Gambar 4.11</b>	Deviasi Standar Kelompok Periode April 2017 – Februari 2018.....	39

<b>Gambar 4.12</b>	Nilai <i>Silhouette Width</i> 8 Klaster Periode Mei 17 – Feb 18.....	40
<b>Gambar 4.13</b>	Rata-rata Kelompok Periode Mei 2017 – Februari 2018 .....	41
<b>Gambar 4.14</b>	Deviasi Standar Kelompok Periode Mei 2017 – Februari 2018.....	42
<b>Gambar 4.15</b>	Nilai <i>Silhouette Width</i> 10 Klaster Periode Juni 2017 – Februari 2018 .....	43
<b>Gambar 4.16</b>	Rata-rata Kelompok Periode Juni 2017 – Februari 2018 .....	44
<b>Gambar 4.17</b>	Deviasi Standar Kelompok Periode Juni 2017 – Februari 2018.....	45
<b>Gambar 4.18</b>	Nilai <i>Silhouette Width</i> 10 Klaster Periode Juli 17 – Feb 18 .....	46
<b>Gambar 4.19</b>	Rata-rata Kelompok Periode Juli 2017 – Februari 2018 .....	47
<b>Gambar 4.20</b>	Deviasi Standar Kelompok Periode Juli 2017 – Februari 2018.....	48
<b>Gambar 4.21</b>	Nilai <i>Silhouette Width</i> 9 Klaster Periode Agustus 17 – Februari 18.....	49
<b>Gambar 4.22</b>	Rata-rata Kelompok Periode Agustus 2017 – Februari 2018.....	50
<b>Gambar 4.23</b>	Deviasi Standar Kelompok Periode Agustus 2017 Hingga Februari 2018 .....	51
<b>Gambar 4.24</b>	Nilai <i>Silhouette Width</i> 7 Klaster Periode Sep 17 – Feb 18.....	52
<b>Gambar 4.25</b>	Rata-rata Kelompok Periode September 2017 – Februari 2018.....	53
<b>Gambar 4.26</b>	Deviasi Standar Kelompok Periode September 2017 – Februari 2018.....	54
<b>Gambar 4.27</b>	Nilai <i>Silhouette Width</i> 8 Klaster Periode Okt 17 – Feb 18.....	55
<b>Gambar 4.28</b>	Rata-rata Kelompok Periode Oktober 2017 – Februari 2018.....	56

<b>Gambar 4.29</b>	Deviasi Standar Kelompok Periode Oktober 2017 – Februari 2018.....	57
<b>Gambar 4.30</b>	Nilai <i>Silhouette Width</i> 10 Klaster Periode Nov 17–Feb 18 .....	58
<b>Gambar 4.31</b>	Rata-rata Kelompok Periode November 2017 – Februari 2018.....	59
<b>Gambar 4.32</b>	Deviasi Standar Kelompok Periode November 2017 – Februari 2018 .....	60
<b>Gambar 4.33</b>	Nilai <i>Silhouette Width</i> 10 Klaster Periode Des 17–Feb 18.....	61
<b>Gambar 4.34</b>	Rata-rata Kelompok Periode Desember 2017 – Februari 2018.....	62
<b>Gambar 4.35</b>	Deviasi Standar Kelompok Periode Desember 2017 – Februari 2018.....	63
<b>Gambar 4.36</b>	Nilai <i>Silhouette Width</i> 8 Klaster Periode Jan – Feb 18.....	64
<b>Gambar 4.37</b>	Rata-rata Kelompok Periode Januari – Februari 2018 .....	65
<b>Gambar 4.38</b>	Deviasi Standar Kelompok Periode Januari – Februari 2018 .....	66

*(Halaman sengaja dikosongkan)*

## DAFTAR TABEL

<b>Tabel 2.1</b>	Kategori Nilai <i>Silhouette Width</i> .....	15
<b>Tabel 3.1</b>	Variable Penelitian .....	19
<b>Tabel 3.2</b>	Struktur Data .....	20
<b>Tabel 3.3</b>	Struktur Data Uji Deteksi Anomali .....	23
<b>Tabel 4.1</b>	Jumlah Pelanggan Tak Aktif (Rumah Kosong).....	27
<b>Tabel 4.2</b>	Nilai Pseudo-F dan ICD Rate Pada Klaster dan Model Terpilih .....	32
<b>Tabel 4.3</b>	Kelompok Anomali Masing-masing Periode .....	67
<b>Tabel 4.4</b>	Hasil Deteksi Anomali Dengan Algoritma <i>LOF</i> .....	69
<b>Tabel 4.5</b>	Jumlah Pelanggan Terdeteksi Anomali Per-Periode .....	70
<b>Tabel 4.6</b>	Jumlah Pelanggan Terdeteksi Anomali Berdasarkan Frekuensi .....	71
<b>Tabel 4.7</b>	Pelanggan Dengan Frekuensi Tertinggi .....	72
<b>Tabel 4.8</b>	Hasil Deteksi Anomali Menggunakan Metode PDAM Surya Sembada Kota Surabaya .....	73
<b>Tabel 4.9</b>	Karakteristik Pelanggan Terdeteksi Anomali .....	74
<b>Tabel 4.10</b>	Pelanggan Yang Tidak Terdeteksi Anomali Dengan Metode Sederhana PDAM .....	76

*(Halaman sengaja dikosongkan)*

## DAFTAR LAMPIRAN

<b>Lampiran 1</b>	Data Pemakaian Air PDAM Surya Sembada Kota Surabaya Periode Maret 2017-Februari 2018.....	85
<b>Lampiran 2</b>	Data Hasil <i>Pre-Processing</i> Pemakaian Air Pelanggan PDAM Surya Sembada Kota Surabaya .....	86
<b>Lampiran 3</b>	Vektor Pemenang Masing-masing Node Grid.....	97
<b>Lampiran 4</b>	Hasil Pengelompokkan 11 Periode .....	108
<b>Lampiran 5</b>	Nilai <i>Silhouette</i> Masing-masing Observasi 11 Periode.....	109
<b>Lampiran 6</b>	Syntax <i>Kohonen-SOM</i> dan <i>LOF</i> pada R .....	110
<b>Lampiran 7</b>	Syntax Menghitung Pseudo F dan ICDrate di R .....	111
<b>Lampiran 8</b>	Surat Keterangan Data Instansi .....	112

*(Halaman sengaja dikosongkan)*



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

PDAM Surya Sembada Kota Surabaya merupakan salah satu perusahaan air minum yang tergolong cukup besar di Indonesia. Perusahaan ini berdiri sejak tahun 1976 dan dimiliki oleh Pemerintah Kota Surabaya. Hingga tahun 2016 pelanggan PDAM Surya Sembada Kota Surabaya mencapai 547.819. Dengan rincian 502.124 adalah pelanggan perumahan, 1.239 adalah pelanggan pemerintah, 38.089 pelanggan perdagangan, 404 pelanggan industri, 3.794 pelanggan sosial umum, 2.163 pelanggan sosial khusus, dan 6 pelanggan pelabuhan. Sedangkan untuk penduduk yang terlayani hingga tahun 2016 adalah sebanyak 3.127.782 penduduk.

Kerugian yang sering dialami oleh pihak PDAM Surya Sembada Kota Surabaya adalah karena kebocoran pipa, kesalahan pada sistem, dan pencurian air. Kesalahan sistem itu sendiri adalah karena meter pencatat air tidak berfungsi dengan baik, atau meter yang buram se-hingga menyebabkan salah catat. Sedangkan pada tahun 2011, telah terjadi sebanyak 100 kasus pencurian air di Kota Surabaya dan kerugian yang dicapai sebesar Rp. 1,1 miliar. Pada tahun 2012 selama Bulan Januari hingga Juli terjadi sebanyak 525 kasus pencurian air. Artinya jika diambil rata-rata, tiap bulannya terdapat 75 kasus pencurian air. Kerugian PDAM Surya Sembada Kota Surabaya pada tahun itu adalah sebanyak Rp. 200 Miliar. Jaringan PDAM Surya Sembada Kota Surabaya semakin meluas, sehingga kasus pencurian air pun juga meningkat. Tahun 2013 terjadi sebanyak 1.800 kasus dan pada tahun 2014 terjadi sebanyak 4.000 kasus. PDAM Surya Sembada Kota Surabaya sudah melakukan upaya dalam menangani permasalahan tersebut, tapi membutuhkan waktu yang cukup lama.

PDAM Surya Sembada Kota Surabaya menggunakan selisih pemakaian air pada bulan sebelumnya untuk mendeteksi apakah pelanggan melakukan kecurangan atau ada kesalahan sistem dalam meter kubik pelanggan. Selain itu, PDAM Surya Sembada Kota Surabaya juga mengharapkan bantuan dari masyarakat sekitar un-

tuk melaporkan jika terdapat pelanggan yang mencurigakan dalam penggunaan konsumsi air sehari-hari. Deteksi anomali menggunakan selisih masih kurang efektif karena membutuhkan waktu yang sangat lama, belum lagi di-tambah banyakan pengguna jasa PDAM Surya Sembada Kota Surabaya yakni sebanyak 547.819 pelanggan. Pencurian air ditandai dengan tidak bertambahnya meter ukur dari bulan sebelumnya ke bulan berikutnya. Hal pertama yang dilakukan oleh PDAM Surya Sembada Kota Surabaya adalah melakukan investigasi pada meter ukur tersebut, jika rusak langsung diganti. Namun setelah pergantian meter ukur juga masih tidak dapat berfungsi, maka ada indikasi bahwa pelanggan tersebut melakukan pencurian air. Terdapat banyak hal yang menyebabkan meter ukur tidak berfungsi, karena rusak atau terdapat saluran pipa ilegal yang tidak melewati alat ukur tersebut.

Kondisi yang terjadi pada konsumsi ilegal maupun kerusakan pada alat ukur merupakan suatu konsumsi yang tidak normal. Pada awalnya konsumen mempunyai pola normal dengan interval tertentu, tetapi pada suatu waktu terjadi kondisi diluar pola tersebut. Hal ini merupakan kondisi tidak normal atau anomali. Kondisi anomali terhadap konsumsi air PDAM cenderung berkaitan dengan kecurangan, kesalahan pengukuran, konsumsi yang tidak terdeteksi, masalah pada saluran air, dan meteran yang sudah tidak berfungsi. Kondisi anomali yang disebabkan oleh kesalahan teknis disebut kerugian teknis, sedangkan kondisi anomali yang disebabkan oleh kecurangan, pemasangan pipa ilegal termasuk dalam kerugian non-teknis.

Deteksi dan identifikasi konsumsi yang tidak normal telah banyak dilakukan dengan menggunakan metode statistik maupun kecerdasan buatan. Deteksi anomali telah banyak dilakukan pada pengguna kartu kredit yang biasa disebut dengan *fraud detection*. Metode deteksi anomali terdapat beberapa macam salah satunya adalah *supervised learning*, *unsupervised learning*. Deteksi anomali dengan *unsupervised learning*. Selain itu deteksi anomali juga sering dilakukan pada konsumsi listrik dan sistem jaringan komputer. Zakariya (2012) membuat aplikasi pendeteksi anomali pada

pola konsumsi listrik pelanggan Kota Surabaya menggunakan algoritma pengelompokan berbasis densitas. Metode tersebut sangat sensitif terhadap data *noise*, serta sangat algoritma berdasarkan densitas akan menghasilkan model yang tidak cukup baik jika densitas data bervariasi (Dang, 2015). Lapidou, Papageorgiu, Kokkinos, Sahu, Gupta, dan Tassiulas (2015) melakukan penelitian yang berjudul *Exploring Patterns in Water Consumption by Clustering*. Penelitian tersebut mengelompokkan konsumsi air di Negara Yunani. Metode yang digunakan adalah algoritma Kohonen SOM. Penelitian tersebut hanya mengelompokkan saja tidak sampai melakukan deteksi anomali, kesimpulan dari penelitian tersebut adalah algoritma Kohonen SOM sangat baik untuk mengelompokkan objek terutama yang memiliki pola-pola tertentu. Deteksi anomali *unsupervised learning* telah dilakukan menggunakan metode *machine learning* seperti *one class-SVM* dan *Fuzzy C-Means* dengan hasil *one-SVM* lebih bagus untuk mendapatkan pola anomali. Nian, Zhang, Tayal, Coleman, dan Li (2016) juga melakukan deteksi anomali pada pelanggan asuransi dengan menggunakan metode *unsupervised spectral ranking*. Landasan dari metode tersebut adalah menggunakan peringkat *spectral* pada masing-masing pengamatan.

Data yang akan digunakan pada penelitian ini adalah data pelanggan PDAM Surya Sembada Kota Surabaya dalam periode 12 bulan terhitung mulai Bulan Februari. Penelitian yang akan dilakukan pada PDAM Surya Sembada Kota Surabaya kali ini menggunakan pengelompokan dengan algoritma *Self-Organizing Maps* (SOM). Pengelompokan menggunakan algoritma *Kohonen SOM* sangat sesuai, karena algoritma tersebut mengelompokkan berdasarkan pola-pola karakteristik pelanggan air PDAM Surya Sembada Kota Surabaya. Pengelompokan dilakukan pada masing-masing periode, misal periode Maret 2017 – Februari 2018, dan seterusnya. Pengelompokan tersebut berdasarkan rata-rata pemakaian air, pemakaian air maksimal, rata-rata tagihan air, tagihan air maksimal, rata-rata pemakaian air di golongan tertentu,

rata-rata pemakaian air di zona tertentu, rata-rata pemakaian air di sub-zona tertentu dan deviasi pemakaian air. Pengelompokan dilakukan terlebih dahulu, karena jika deteksi anomali dilakukan secara langsung dengan jumlah data yang sangat besar akan menghasilkan model yang tidak cukup baik (Goldstein dan Uchida, 2012). Penentuan kelompok yang optimal menggunakan nilai *Pseudo F*. Setelah ditemukan jumlah kelompok yang optimal, deteksi anomali dilakukan pada kelompok dengan nilai *silhouette width* yang paling rendah karena berarti kelompok tersebut mempunyai sebaran data yang sangat besar. Kemudian deteksi anomali dilakukan dengan menggunakan *Local Outlier Factor* (LOC), yang merupakan metode yang cukup baik dalam mendeteksi *outlier* baik secara local maupun global (Goldstein dan Uchida 2012).

## 1.2 Rumusan Masalah

Kerugian dalam masalah pendistribusian air dapat merugikan berbagai pihak seperti PDAM Surya Sembada Kota Surabaya, pemerintah kota dan masyarakat. Kerugian tersebut dapat disebabkan oleh kesalahan teknis atau kerusakan pada catat meter bahkan pencurian air. Berdasarkan latar belakang yang telah dipaparkan dirumuskan suatu rumusan permasalahan untuk mengetahui anomali konsumsi air di Kota Surabaya menggunakan metode *Kohonen SOM* dan *Local Outlier Factor*. Anomali adalah data yang berbeda dari pengamatan pada umumnya dan mempunyai arti. *Kohonen SOM* digunakan untuk mengelompokkan pelanggan PDAM Surya Sembada Kota Surabaya dan kemudian dianalisa menggunakan *Local Outlier Factor*. PDAM Surya Sembada Kota Surabaya sendiri juga mempunyai metode sederhana dalam mendeteksi anomali, yaitu dengan menghitung selisih pemakaian bulan saat ini dan bulan sebelumnya. Rumusan masalah selanjutnya adalah mendapatkan karakteristik pelanggan yang terdeteksi anomali menggunakan metode *Kohonen SOM* dan *Local Outlier Factor*.

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, penelitian ini memiliki beberapa tujuan untuk dicapai yaitu sebagai berikut:

1. Mendapatkan hasil deteksi anomali pada konsumsi air pelanggan PDAM Surya Sembada Kota Surabaya dengan menggunakan algoritma *Local Outlier Factor*.
2. Mendapatkan karakteristik pola konsumsi pelanggan PDAM Surya Sembada Kota Surabaya yang terdeteksi sebagai anomali.

### 1.4 Manfaat Penelitian

Hasil penelitian ini bermanfaat terhadap PDAM Surya Sembada Kota Surabaya, khususnya dalam meminimumkan kerugian yang dialami PDAM Surya Sembada Kota Surabaya terkait *water loss* yang disebabkan oleh konsumsi tidak wajar yang dikarenakan kesalahan sistem atau pencurian air. Selain itu, penelitian ini dapat digunakan sebagai tambahan informasi untuk mempercepat untuk mengetahui bagaimana karakteristik pelanggan PDAM Surya Sembada Kota Surabaya yang terdeteksi sebagai anomali.

### 1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut:

1. Data yang akan digunakan pada penelitian adalah pada periode Maret 2017 hingga Februari 2018.
2. Pelanggan dalam penelitian ini adalah pelanggan yang menggunakan jasa PDAM Surya Sembada Kota Surabaya dalam periode Maret 2017 hingga Februari 2018.
3. Pelanggan yang digunakan pada penelitian ini adalah pelanggan rumah tangga saja, untuk perusahaan dan sejenisnya tidak termasuk.
4. Jika terdapat pelanggan yang tiba-tiba berhenti berlangganan atau mulai berlangganan pada periode tersebut, akan dianggap sebagai *missing values* dan tidak disertakan dalam analisa.

5. Analisa hanya menggunakan data pelanggan aktif, yang berarti pemakaian air di bangunan/rumah tersebut tidak kosong selama 3 hingga 12 bulan berturut-turut.

## **BAB II**

### **TINJAUAN PUSTAKA**

Pada bagian ini membahas mengenai beberapa kajian pustaka yang digunakan untuk menyelesaikan permasalahan mengenai anomali pada konsumsi air pelanggan PDAM Surya Sembada Kota Surabaya. Kajian pustaka tersebut diantaranya deteksi anomali, pengelompokan *Kohonen Self-Organizing Maps*, deteksi *outlier*, anomali pada konsumsi air, dan penelitian sebelumnya.

#### **2.1 Deteksi Anomali**

Deteksi anomali menggunakan pendekatan berbasis data pada masa lalu. Data anomali adalah sebuah data yang tidak normal berdasarkan pola-pola sebelumnya. Anomali dapat dikategorikan kedalam beberapa kategori sebagai berikut:

1. Anomali titik, sebuah data yang memiliki nilai jauh berbeda dibanding-kan data lainnya.
2. Anomali kontekstual, jenis anomali ini biasanya terjadi dalam deret waktu.
3. Anomali kolektif, kumpulan contoh suatu data yang digunakan atau dapat membantu dalam mendeteksi anomali

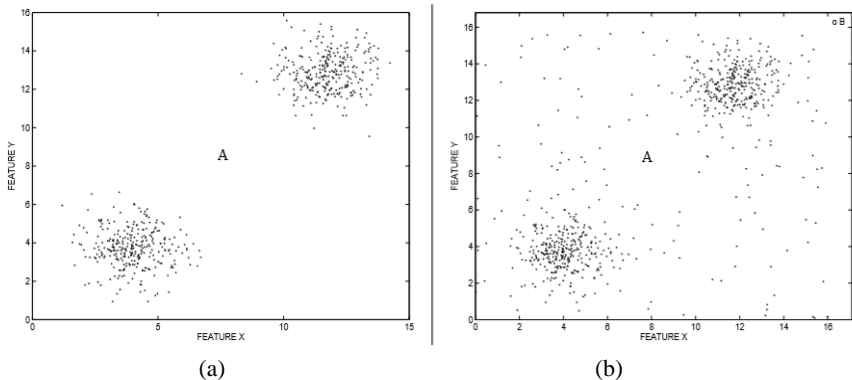
Deteksi anomali serupa dengan halnya menghilangkan *noise* pada suatu data. Anomali juga dapat disebut *outlier*, yaitu objek data yang berbeda secara signifikan dibandingkan dengan objek yang lain.

Terdapat beberapa cara dalam mendeteksi anomali, diantaranya adalah *supervised learning*, *unsupervised learning*, deteksi *outlier*, *semi-supervised learning* atau dengan kata lain pendekatan statistik dan metode berbasis *machine learning*. Pendekatan statistika dapat menggunakan rata-rata berjalan untuk melihat pola berdasarkan data masa lalu. Sedangkan *supervised learning* adalah melakukan prediksi atau membentuk model berdasarkan *training* data yang telah memiliki label. Sedangkan *unsupervised* melakukan deteksi anomali dengan melakukan klastering, atau pengelompokan. Dalam pengaplikasiannya *unsupervised learning* biasa digunakan untuk mendeteksi kecurangan pada pengguna kartu kredit.

Hasil dari deteksi outlier atau anomali dapat di golongkan kedalam dua jenis sebagai berikut (Aggarwal, 2013):

1. Nilai anomali/outlier, sebagian algoritma menghasilkan suatu nilai yang akan digunakan untuk menentukan apakah observasi tersebut termasuk kedalam jenis anomali atau tidak.
2. Label biner, sebagian algoritma menghasilkan suatu label yang bersifat biner. Yaitu normal dan anomali. Tipe kedua ini sangat dibutuhkan dalam aplikasi sehari-hari.

Keputusan dalam menentukan apakah deviasi dari suatu titik tersebut dapat dianggap sebagai anomali merupakan hal yang sangat subjektif. Karena pasti akan terdapat *noise* yang dapat menyebabkan anomali/outlier sulit terdeteksi. Pada aplikasi yang nyata terdapat banyak *noise* yang terbentuk, namun *noise* tidak mempunyai arti dalam data tersebut. Hal yang menarik untuk dianalisis adalah deviasi data yang berbeda secara signifikan. Seperti yang dijelaskan pada Gambar 2.1, gambar tersebut menunjukkan bahwa pada Gambar 2.1 (a) menjelaskan titik anomali yang bernama A berbeda signifikan dengan data yang lain. Anomali yang terlihat sangat jelas karena hasil kelompok yang terbentuk memiliki



(Sumber: *Outlier Analysis*, Aggarwal 2013)

**Gambar 2.1** Perbedaan Anomali dan *Noise*

keragaman yang kecil. Hal ini berbeda dengan titik A pada Gambar 2.1 (b) yang terlihat lebih subjektif dan juga terlihat pada daerah data yang jarang serta terdapat banyak *noise* sehingga titik anomali



sangat subjektif sekali untuk ditentukan. Oleh karena itu, *outlier* berkenaan dengan suatu titik data dimana dapat dinyatakan sebagai anomali/*noise* dengan anomali merupakan kondisi khusus yang mempunyai arti pada *outlier* yang sangat menarik untuk dianalisa (Aggarwal, 2013).

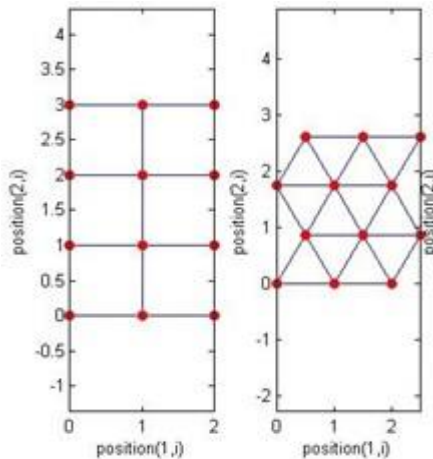
Pada metode *unsupervised*, dimana tidak tersedia keterangan pada data yang menunjukkan kelas *outlier*, *noise* merepresentasikan batas yang berkaitan antara data normal dengan anomali yang sesungguhnya. *Noise* sering dimodelkan sebagai bentuk lemah dari *outlier* yang tidak selalu memenuhi kriteria yang dapat dikategorikan sebagai anomali atau dengan kata lain tidak mempunyai arti khusus. Batas perbedaan antara region tersebut tidak didefinisikan dengan pasti, karena masing-masing data mempunyai keragaman yang tidak absolut. Oleh karena itu, anomali mempunyai nilai *outlier* yang lebih tinggi dibandingkan dengan *noise* dan mempunyai arti khusus, tetapi tidak dapat menjadi faktor penentu antara kedua kategori tersebut dalam hal definisi (Aggarwal, 2013).

## 2.2 Kohonen Self-Organizing Maps (SOM)

Metode Kohonen *Self-Organizing Maps* adalah salah satu bagian khusus dari *neural network*. *Self-Organizing Maps* mempunyai tujuan untuk mengkonversi input dengan data yang berdimensi besar menjadi pemetaan diskrit yang sederhana berdimensi rendah (Haykin, 1990). Hal tersebut sangat tepat untuk digunakan menganalisa kelompok yang didalamnya terdapat pola tersembunyi diantara objek yang dianalisa (Larose, 2005).

Seperti kebanyakan jaringan syaraf tiruan, SOM beroperasi dalam dua mode, yaitu *training* dan *mapping*. *Training* membentuk *map* dengan menggunakan vektor awal sementara *mapping* mengklasifikasikan vektor masukan baru. Node *output* atau neuron pada jaringan Kohonen diletakkan dalam 2 dimensi yang bentuk atau topologinya dapat diatur. Topologi yang berbeda akan menghasilkan neuron dan neuron pemenang yang berbeda pula sehingga bobot yang diubah akan berbeda pula. Terdapat tiga

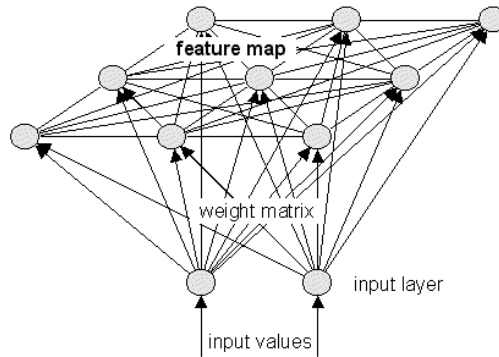
macam topologi yang dapat dibentuk namun pada penelitian ini hanya menggunakan 2 macam topologi, yaitu *gridtop*, dan *hextop*, dapat dilihat pada Gambar 2.2 berikut.



(Sumber: intechopen.com)

**Gambar 2.2** Jenis Topologi *Gridtop* dan *Hextop*

Gambar 2.2 menunjukkan bahwa bagian kiri merupakan topologi *gridtop*, dan bagian kanan adalah topologi *hextop*. Penggunaan topologi tersebut akan menghasilkan vektor bobot dan hasil pemetaan yang berbeda. Tidak dapat ditentukan manakah topologi yang akan menghasilkan kelompok dengan nilai kesalahan yang paling baik. Perbedaan dari kedua jenis topologi tersebut adalah penyusunan node *output*. Jika menggunakan topologi *gridtop*, maka penyusunan node *output* adalah berjalan dengan cara persegi, sedangkan pada topologi *hextop* disusun secara *hexagon* atau piramida. Pada penelitian ini topologi yang akan digunakan untuk mengelompokkan data adalah yang menghasilkan nilai *Pseudo-F* yang terbesar. Sehingga penggunaan dari topologi tersebut akan menghasilkan vektor bobot dan hasil pemetaan yang berbeda. Bentuk dari jaringan SOM dapat dilihat pada Gambar 2.3.



(Sumber: <http://www.nnwj.de/kohonen-feature-map.html>)

**Gambar 2.3** Jaringan Kohonen SOM

Secara singkat, SOM membutuhkan bobot dan *input values* pada input layer. Sedangkan bobot tersebut terletak pada masing-masing neuron. Vektor bobot berfungsi sebagai vektor yang mewakili pola-pola vektor masukan atau *input values*. Selama proses pembentukan atau *mapping*, vektor bobot yang mempunyai kemiripan dengan vektor masukan dinyatakan sebagai neuron pemenang. Kemudian neuron pemenang tersebut dan neuron-neuron disekitarnya akan memperbarui vektor bobot masing-masing. Berikut adalah algoritma dari Kohonen *Self-Organizing Maps*.

1. Tahap inisiasi. Pada tahap ini dilakukan pemilihan nilai acak sebagai vektor bobot awal  $w_j(0)$  pada masing-masing neuron. Akan lebih baik jika mengetahui bobot yang tepat untuk mendapatkan hasil yang terbaik. Nilai bobot  $w_j$  berbeda-beda pada masing-masing neuron.
2. Memasukkan vektor  $x(n)$  pada *input values*.
3. Menemukan neuron yang paling mirip atau neuron pemenang pada setiap langkah  $n$  dengan menggunakan jarak euclidean minimum antara vektor masukan dan vektor bobot di masing-masing neuron.

$$D(\mathbf{x}, \mathbf{w}_j) = \sqrt{\sum_{i=1}^n (\mathbf{x}_i - \mathbf{w}_j)^2} \quad (2.1)$$

Keterangan,

$\mathbf{x}_i$  : Vektor input, dengan  $i = 1, 2, 3, \dots, n$

$\mathbf{w}_j$  : Vektor bobot, dengan  $j = 1, 2, 3, \dots, l$

$\mathbf{x}_i$  adalah vektor input pada *input values* yang berarti adalah data asli. Sedangkan  $\mathbf{w}_j$  adalah vektor bobot pada neuron-neuron ke- $j$ . Dengan persamaan 2.1 didapatkan neuron pemenang yaitu yang memiliki jarak paling minimum antara vektor input dengan vektor bobot.

4. Menyesuaikan vektor bobot berdasarkan semua neuron dengan menggunakan persamaan 2.2.

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \alpha(n)h_{j,i(x)}(\mathbf{x}(n) - \mathbf{w}_j(n)) \quad (2.2)$$

Keterangan,

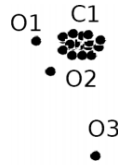
$\alpha$  : Parameter tingkat pembelajaran pada iterasi ke- $n$ .

Dengan  $\alpha$  adalah parameter tingkat pembelajaran, dimana dapat digunakan sebagai berhentinya iterasi dengan menentukan batas pada nilai alpha. Kemudian  $h_{j,i(x)}$  adalah fungsi neuron tetangga disekitar neuron pemenang.

5. Melakukan langkah 2 hingga 4 sampai iterasi terhenti. Iterasi dapat terhenti jika nilai error sudah memenuhi kriteria, atau dengan menentukan jumlah iterasi, atau dengan menentukan kriteria pada nilai alpha.

### 2.3 Local Outlier Factor

*Local Outlier Factor* adalah algoritma deteksi *outlier* yang biasanya digunakan dalam deteksi anomali berdasarkan perbandingan kepadatan lokal dengan data disekitarnya. Dalam *outlier*, istilah *local outlier* adalah untuk melambangkan *outlier* yang sangat dekat dengan kelompok tertentu. Sebagai contoh dapat dilihat pada Gambar 2.4.



(Sumber: home.cse.ust.hk)

**Gambar 2.3** Ilustrasi Outlier

Gambar 2.4 menjelaskan bahwa  $O_1$  merupakan *outlier* lokal untuk kelompok  $C_1$ , begitu juga dengan  $O_2$ . Sedangkan  $O_3$  merupakan global *outlier* untuk  $C_1$ . Metode ini sangat efektif untuk mendeteksi anomali yang bersifat lokal dan global. Suatu pengamatan dikatakan sebagai *global outlier* ketika nilai *outlier* tersebut lebih besar secara signifikan terhadap nilai *outlier*  $K$  pengamatan terdekatnya. Sedangkan suatu pengamatan dikatakan sebagai *local outlier* ketika nilai *outlier* relative lebih besar dari rata-rata nilai *outlier*  $K$  pengamatan terdekatnya. Dalam menentukan *threshold* tidak terdapat ketentuan karena setiap kasus, dan data memiliki deviasi standar yang berbeda (Dang, Micenkova, Assent, & Ng, 2013). Sehingga ketika anomali lokal terdeteksi, global juga pastinya akan terdeteksi. Metode ini mempunyai tahapan-tahapan, yaitu menentukan nilai *K-Nearest Neighbor*, kemudian menghitung LRD, dan terakhir mendapatkan nilai LOF. Berikut adalah algoritmanya.

1. Menentukan nilai  $K$ , nilai  $K$  adalah jumlah tetangga terdekat yang akan dihitung jaraknya. Misal diberikan nilai  $K$  sebesar 5, maka sebanyak 5 tetangga terdekat pada masing-masing data akan dihitung jaraknya.
2. Menghitung jarak pada masing-masing data, jarak yang dapat digunakan seperti *euclidean* dan *manhattan*.
3. Menghitung jarak terhadap banyaknya pengamatan berdasarkan nilai  $K$ . Dengan kata lain menghitung jarak antara pengamatan dengan  $K$  pengamatan yang terdekat berada disekitarnya. Jika jarak terhadap  $K$  pengamatan mempunyai nilai lebih kecil atau sama dengan jarak

terdekatnya, maka hitung berapa jumlah pengamatan tersebut dan notasikan sebagai  $N_k(o)$ .

4. Menghitung nilai *local reachability density* dinotasikan dalam  $lrd_k(o)$ , rumus yang digunakan dapat dilihat pada persamaan 2.3

$$lrd_k(o) = \frac{\|N_k(o)\|}{\sum_{o' \in N_k(o)} reachdist_k(o' \leftarrow o)} \quad (2.3)$$

Keterangan,

$N_k(o)$  :  $k$  objek yang terdekat pada objek  $o$

Persamaan 2.3 menunjukkan  $lrd_k(o)$  atau nilai jangkauan kepadatan suatu objek pada objek tertentu.  $\|N_k(o)\|$  menunjukkan jumlah objek  $o'$  yang terdekat pada objek  $o$  setelah objek yang terdekat pertama sebanyak  $k$  objek dengan nilai  $k$  yang ditentukan. Sedangkan  $reachdist_k(o' \leftarrow o)$  nilai maksimal atau tertinggi antara jarak objek  $o$  pada  $k$  objek disekitarnya dengan jarak objek  $o$  ke  $o'$ .

5. Menghitung nilai *local outlier factor* pada masing-masing pengamatan dapat dinotasikan sebagai  $LOF_k(o)$ , rumus yang digunakan dapat dilihat pada persamaan 2.4.

$$LOF_k = \frac{\sum_{o' \in N_k(o)} \frac{lrd_k(o')}{lrd_k(o)}}{\|N_k(o)\|} \quad (2.4)$$

6. Mengurutkan nilai  $LOF$  dari tertinggi ke terendah, kemudian evaluasi apakah objek ke- $k$  merupakan anomali. Evaluasi tersebut dapat menggunakan nilai threshold yang ditentukan atau jika nilai tersebut jauh lebih besar dari nilai 1.

## 2.4 Koefisien *Silhouette*

Koefisien *Silhouette* dapat digunakan untuk mempelajari jarak pemisahan antara klaster yang dihasilkan. Koefisien ini memberikan gambaran tentang seberapa baik setiap objek berada didalam klasternya. Nilai *Silhouette* adalah ukuran seberapa mirip

sebuah objek dengan klasternya sendiri (kohesi) dibandingkan dengan kelompok lain. Nilai *Silhouette* berkisar antara -1 sampai +1, dengan nilai tinggi menunjukkan bahwa objek tersebut sesuai dengan anggota didalam kelompoknya dan kurang sesuai dengan anggota kelompok lainnya. Sehingga yang diharapkan adalah nilai *Silhouette* yang tinggi. Nilai *Silhouette* dapat dihitung dengan metrik jarak, seperti jarak *euclidean* atau jarak *manhatan*.

$$s(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (2.5)$$

Dengan,

- $a_i$  : Nilai rata-rata jarak untuk semua objek didalam klasternya.  
 $b_i$  : Nilai rata-rata jarak semua objek pada klaster yang berbeda.  
 $s_i$  : Nilai koefisien *Silhouette*.

Sehingga diharapkan adalah nilai  $a_i$  seminim mungkin, artinya jarak antar objek dalam satu kelompok sangat kecil, dan nilai  $b_i$  semaksimal mungkin, yang berarti keragaman antar kelompok sangat tinggi. Sehingga didapatkan nilai  $s_i$  mendekati 1 (Amorim & Henning, 2015)

Setelah mendapatkan nilai *silhouette* pada objek tiap kelompok, dapat diketahui nilai *silhouette width* untuk masing-masing kelompok. Nilai tersebut dapat diketahui dengan menghitung nilai rata-rata *silhouette* semua objek yang ada dalam kelompok tersebut. Berikut adalah interpretasi *silhouette width*.

**Tabel 2.1** Kategori Nilai *Silhouette Width*

<i>Silhouette Width</i>	Intepretasi
0,71 – 1	Kelompok yang kuat
0,51 – 0,7	Kelompok yang telah layak atau sesuai
0,26 – 0,5	Kelompok yang lemah
$\leq 0,25$	Tidak dapat dikatakan sebagai kelompok

## 2.5 Pseudo-F Statistics

*Pseudo-F Statistics* digunakan untuk menentukan jumlah kelompok optimum dalam pengelompokkan. Berikut adalah

persamaan yang digunakan untuk menghitung nilai *Pseudo F-Statistics*.

$$Pseudo\ F-Statistics = \frac{\left( \frac{R^2}{n_c - 1} \right)}{\left( \frac{1 - R^2}{n - n_c} \right)} \quad (2.6)$$

Dengan:

$$R^2 = \frac{SST - SSE}{SST} \quad (2.7)$$

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (V_{ij}^k - \bar{V}^k)^2 \quad (2.8)$$

$$SSE = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (V_{ij}^k - \bar{V}_i^k)^2 \quad (2.9)$$

Keterangan:

*SST* : Total jumlah dari kuadrat jarak titik data terhadap rata-rata keseluruhan

*SSE* : Total jumlah dari kuadrat jarak titik data terhadap rata-rata kelompoknya

$R^2$  : Nilai kebaikan model.

$n_c$  : Jumlah klaster yang terbentuk atau banyaknya kelompok yang terbentuk

$n$  : Banyaknya sampel.

$n_i$  : Banyaknya observasi pada kelompok ke- $i$ .

$n_v$  : Banyaknya variabel yang digunakan.

$V_{ij}^k$  : Nilai pada variabel ke- $k$  observasi ke- $j$  dari kelompok ke- $i$

$\bar{V}^k$  : Rata-rata seluruh sampel pada variabel ke- $k$ .

$\bar{V}_i^k$  : Rata-rata seluruh sampel pada kelompok ke- $i$  dan variabel ke- $k$

Nilai *Pseudo F-statistics* tertinggi menunjukkan bahwa jumlah kelompok yang digunakan untuk memartisi data telah optimal. Dimana keragaman dalam kelompok tersebut sangat homogen atau kecil, sedangkan antar kelompok keragamannya sangat besar atau heterogen.



## 2.6 Anomali Pada Konsumsi Air

Anomali pada konsumsi air biasanya disebabkan oleh beberapa faktor, yaitu jaringan pipa, struktur tanah, alat meteran, dan pengguna. Anomali pada konsumsi air dapat merugikan berbagai pihak, terutama perusahaan penyedia air. Karena anomali yang dibiarkan pada konsumsi air akan menyebabkan kerugian finansial bagi perusahaan. Anomali merupakan suatu kejadian dimana menghasilkan pola yang tidak sesuai dengan pola pada data masa lalu. Pada konsumsi air, hal ini dapat disebabkan karena catat meter yang rusak, terjadi kebocoran pipa, atau memang terdapat kecurangan pada pelanggan tersebut.

## 2.7 Penelitian Sebelumnya

Deteksi anomali telah banyak sejak tahun 2000-an. Metode yang digunakan terdapat 3 jenis, pendekatan statistika, *supervised learning*, dan *unsupervised learning*. Metode pendekatan statistika adalah menggunakan rata-rata berjalan (*moving average*), sedangkan *supervised learning* adalah klasifikasi seperti menggunakan metode SVM dan NBC. Deteksi anomali pada umumnya banyak digunakan untuk mendeteksi kecurangan pada transaksi kartu kredit, permasalahan pada jaringan komputer, atau kesalahan pada sistem tertentu.

Angelos, Saavedra, Cortes, dan Souza (2011) melakukan penelitian deteksi dan identifikasi ketidaknormalan pada konsumsi pelanggan dalam distribusi listrik. Metode yang digunakan adalah *Fuzzy C-Means*, variabel yang digunakan dalam penelitian tersebut adalah  $M_6$  yang merepresentasikan rata-rata pemakaian listrik pelanggan, kemudian  $Max_6$  yang merepresentasikan maksimal pemakaian pelanggan, dan  $Dev_6$  Deviasi standar pemakaian listrik pelanggan. Semua variabel tersebut dalam kurun waktu tertentu. Penelitian tersebut menggunakan data pada waktu Juli hingga November 2008. *Fuzzy C-Means* sangat bagus pada kasus ini, karena variabel yang digunakan tidak begitu banyak, selain itu tidak terdapat asumsi yang harus dipenuhi. Lapidou, Papageorgiu, Kokkinos, Sahu, Gupta, dan Tassioulas (2015) melakukan penelitian pengelompokkan berdasarkan pola konsumsi air di

Negara Yunanai. Metode yang digunakan adalah algoritma Kohonen SOM, variabel yang digunakan diantaranya rata-rata penggunaan air dalam 4 bulan, rasio rata-rata pemakaian dengan maksimal pemakaian, rasio rata-rata pemakaian dalam kuartal pertama dengan rata-rata pemakaian dalam kuartal ketiga, dan seterusnya. Hasil yang didapatkan adalah metode ini sangat cocok digunakan dalam kasus ini, karena algoritma SOM adalah pengelompokkan dengan pendekatan *neural network*, dan dapat melihat pola-pola tertentu terhadap data.

## **BAB III**

### **METODOLOGI PENELITIAN**

Bagian ini membahas mengenai metode penelitian yang digunakan untuk menyelesaikan permasalahan deteksi anomali pada konsumsi air pelanggan PDAM Surya Sembada Kota Surabaya yang meliputi sumber data, variabel penelitian, struktur data, dan langkah analisis.

#### **3.1 Sumber Data**

Data yang digunakan pada penelitian kali ini adalah berasal dari PDAM Surya Sembada Kota Surabaya. Data tersebut merupakan data pemakaian air masing-masing pelanggan di Kota Surabaya. Data bersifat sekunder yang dicatat oleh petugas dan kemudian diinput oleh bagian pemakaian. Periode waktu yang digunakan adalah Maret 2017 hingga Februari 2018. Dalam PDAM Surya Sembada Kota Surabaya terdapat 6 zona dan 42 sub zona. Penelitian ini menggunakan semua zona dan sub-zona. Serta dalam PDAM pelanggan dibagi kedalam berdasarkan golongan/kelas.

#### **3.2 Variabel Penelitian**

Variabel penelitian yang digunakan pada penelitian ini adalah sebagai berikut:

**Tabel 3.1 Variable Penelitian**

<b>Variabel</b>	<b>Keterangan</b>	<b>Skala</b>
$X_1$	Rata-rata konsumsi air pelanggan ( $m^3$ )	Rasio
$X_2$	Konsumsi maksimum pelanggan ( $m^3$ )	Rasio
$X_3$	Rata-rata konsumsi air berdasarkan golongan/kelas pelanggan ( $m^3$ )	Rasio
$X_4$	Rata-rata konsumsi air berdasarkan zona ( $m^3$ )	Rasio
$X_5$	Rata-rata konsumsi air berdasarkan sub-zona ( $m^3$ )	Rasio
$X_6$	Deviasi standar konsumsi pelanggan PDAM Surya Sembada Kota Surabaya ( $m^3$ )	Rasio

Penjelasan terkait variabel-variabel yang digunakan pada penelitian ini adalah sebagai berikut:

1. Variabel  $X_1$  merepresentasikan rata-rata konsumsi air masing-masing pelanggan selama selang waktu tertentu sebagai estimasi pola khusus konsumsi air pelanggan PDAM Surya Sembada Kota Surabaya.
2. Variabel  $X_2$  merepresentasikan pemakaian air tertinggi oleh pelanggan pada selang waktu tertentu.
3. Variabel  $X_3$  merepresentasikan rata-rata konsumsi air berdasarkan golongan pelanggan atau kelas pelanggan pada selang waktu tertentu. Digunakan untuk mengetahui pola konsumsi pelanggan pada masing-masing kelas/golongan. Golongan yang digunakan adalah golongan rumah tangga yang terbagi kedalam 5 kelas, yaitu 2A II, 3A, 3C II, 4A, dan 4B II.
4. Variabel  $X_4$  merepresentasikan rata-rata konsumsi air di zona tertentu pada periode waktu tertentu. Digunakan untuk mengetahui pola konsumsi pelanggan pada zona tertentu. PDAM Surya Sembada Kota Surabaya mempunyai 6 zona. Pembagian zona PDAM didasarkan pada instalasi, yaitu Instalasi Ngagel I, Instalasi Ngagel II, Instalasi Ngagel III, Instalasi Kayoon, dan Instalasi Karangpilang. Sedangkan satu zona sisanya, yaitu zona luar daerah Surabaya.
5. Variable  $X_5$  merepresentasikan rata-rata konsumsi pelanggan di sub-zona tertentu pada periode tertentu. Digunakan untuk mengetahui pola konsumsi pelanggan di sub-zona tertentu.
6. Variabel  $X_6$  merepresentasikan deviasi standar selama periode waktu tertentu untuk melihat pola keragaman pelanggan.

### 3.3 Struktur Data

Struktur data yang digunakan pada penelitian ini akan disajikan pada Tabel 3.2

**Tabel 3. 2 Struktur Data**

Nomor	$X_1$	$X_2$	...	$X_6$
1	$X_{1,1}$	$X_{2,1}$	...	$X_{6,1}$
2	$X_{1,2}$	$X_{2,2}$	...	$X_{6,2}$
...	...	...	...	...
$n$	$X_{1,n}$	$X_{2,n}$		$X_{6,n}$

Struktur data yang terbentuk adalah pada Tabel 3.2 adalah pada selang periode tertentu. Nomor menjelaskan banyaknya pengamatan hingga  $n$  pelanggan. Sedangkan  $X_1$  hingga  $X_6$  berdasarkan pada Tabel 3.1 dan memiliki jumlah sebanyak  $n$  pelanggan. Pelanggan yang digunakan adalah pelanggan rumah tangga. Data yang digunakan pada penelitian ini dilakukan selama Maret 2017 hingga Februari 2018. Sehingga akan terbentuk sebanyak 12 tabel yang serupa dengan Tabel 3.2, pada masing-masing selang periode.

### 3.4 Langkah Analisis

Langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut,

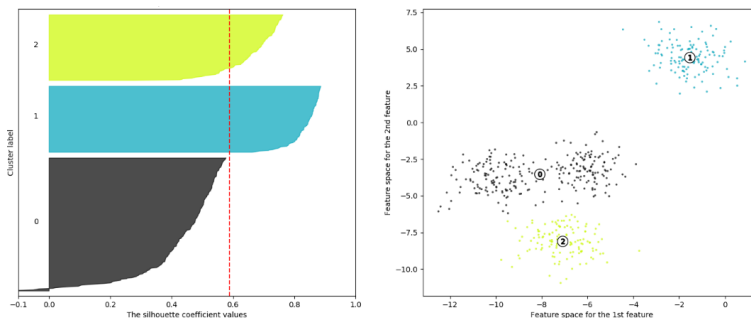
1. Mengumpulkan data yang berasal dari divisi pemakaian PDAM Surya Sembada Kota Surabaya. Data yang dikumpulkan berisi tentang penggunaan konsumsi air, zona pelanggan, golongan pelanggan, subzona pelanggan, dan alamat. Data berasal dari periode Bulan Maret 2017 hingga Februari 2018.
2. Melakukan *pre-processing*.  
 Pada tahap ini akan dilakukan proses untuk menyiapkan data yang akan dianalisa. Berikut adalah beberapa hal yang dilakukan selama *pre-processing*.
  - a. Memilah pelanggan yang mengonsumsi air selama periode penelitian pada kedalam 12 periode.
  - b. Mengatasi *missing value* dengan cara menghilangkan data yang hilang atau tidak diikuti pada analisa. *Missing value* dapat terjadi ketika terdapat pelanggan yang baru berlangganan atau berhenti berlangganan ditengah-tengah periode penelitian.
  - c. Mendapatkan pelanggan aktif dengan syarat pelanggan yang melakukan konsumsi air selama setidaknya 10 bulan.
  - d. Data pelanggan yang didapatkan pada langkah 1 diolah sehingga didapatkan variabel seperti pada Tabel 3.1 dan struktur data pada Tabel 3.2.

3. Melakukan pengelompokkan dan mendapatkan jumlah kelompok yang optimal.

Pada tahap ini dilakukan pengelompokkan menggunakan algoritma *Kohonen SOM* algoritma pada subbab 2.2. Kelompok yang akan dibentuk sebanyak 2 hingga 7 kelompok. Selanjutnya menghitung nilai *Pseudo F* pada masing-masing kelompok untuk mendapatkan jumlah kelompok yang optimal.

4. Mendapatkan kelompok anomali.

Setelah dilakukan jumlah kelompok yang optimal, selanjutnya mendapatkan kelompok yang mempunyai sebaran data tidak normal. Cara yang digunakan adalah dengan melihat nilai *silhouette width* pada masing-masing kelompok di tiap periode. Nilai *silhouette width* yang rendah merepresentasikan bahwa sebaran data sangat luas dan terdapat data-data yang tidak normal, seperti yang disajikan pada Gambar 3.1.



(Sumber: scikit-learn.org)

**Gambar 3.1** Ilustrasi Nilai Silhouette Width Terhadap Sebaran Data

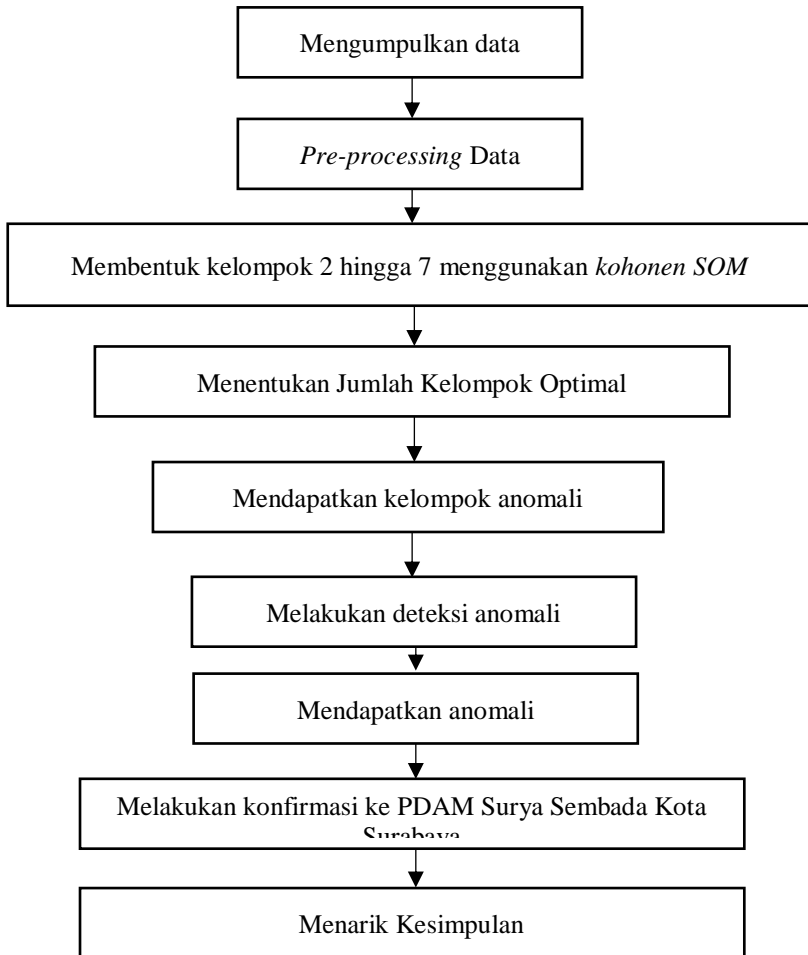
5. Melakukan deteksi anomali

Deteksi anomali dilakukan pada masing-masing periode berdasarkan kelompok yang terpilih pada langkah 5. Deteksi anomali menggunakan algoritma *Local Outlier Factor* (algoritma pada subbab 2.3) dalam kurun waktu satu tahun. Berikut adalah struktur data tabel hasil deteksi anomali pada periode yang berbeda-beda.

**Tabel 3.3** Struktur Data Uji Deteksi Anomali

<b>Periode</b>	<b>Jumlah Pelanggan yang Terdeteksi</b>	<b>Nomor Observasi Pelanggan</b>
Maret 2017 – Februari 2018	...	...
April 2017 – Februari 2018	...	...
...	...	...
Desember 2017 – Februari 2018	...	...
Januari 2018 – Februari 2018	...	...

7. Mendapatkan anomali yang sering muncul pada tiap periode. Berdasarkan hasil deteksi tersebut, dapat dihitung frekuensi kemunculan suatu pelanggan dalam kurun waktu 1 tahun. Jika pada jumlah frekuensi tertentu terdapat pelanggan terdeteksi sebagai anomali, maka dapat disimpulkan bahwa pelanggan tersebut merupakan anomali atau terdapat konsumsi yang tidak normal.
8. Melakukan konfirmasi apakah anomali yang terdeteksi dengan menggunakan metode penelitian merupakan anomali yang diduga oleh PDAM Surya Sembada Kota Surabaya. Kemudian melihat karakteristik pelanggan yang terdeteksi sebagai anomali pada data pelanggan PDAM Surya Sembada Kota Surabaya.
9. Membuat Kesimpulan  
Secara ringkas langkah analisis disajikan dalam diagram alir pada Gambar 3.2



**Gambar 3.2** Diagram Alir Penelitian



## **BAB IV**

### **ANALISIS DAN PEMBAHASAN**

Bagian ini akan memaparkan mengenai analisis dan pembahasan tentang deteksi anomali pada konsumsi air pelanggan PDAM Surya Sembada Kota Surabaya periode Maret 2017 hingga Februari 2018. Pembahasan meliputi pengelompokkan, pemilihan kelompok anomali, dan deteksi anomali. Metode yang digunakan untuk membentuk kelompok adalah *Kohonen Self-Organizing Maps*, pemilihan kelompok anomali menggunakan nilai *silhouette*, dan deteksi anomali menggunakan *Local Outlier Factor*.

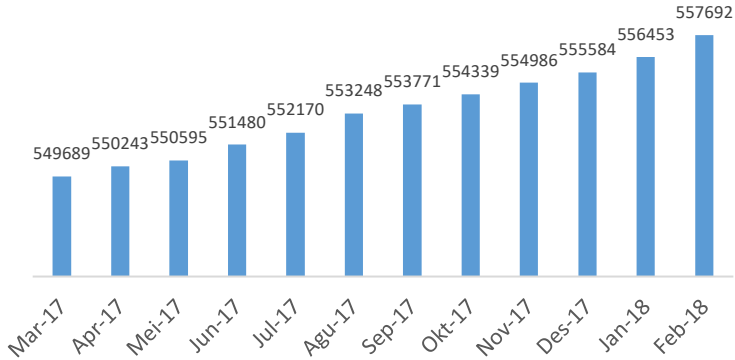
Pengelompokkan tersebut dilakukan untuk mengetahui pola pelanggan dengan konsumsi yang serupa. Sebelum melakukan pengelompokkan terlebih dahulu mendapatkan variabel-variabel yang akan digunakan dalam membentuk kelompok, yaitu rata-rata pemakaian air, maksimum pemakaian air, rata-rata pemakaian air zona, rata-rata pemakaian air sub-zona, rata-rata pemakaian air golongan tarif, dan deviasi standar. Setelah dilakukan pengelompokkan, mendapatkan nilai *silhouette* pada masing-masing kelompok yang terbentuk di masing-masing periode. Pemilihan kelompok anomali dengan memperhatikan nilai *silhouette* tidak memenuhi kriteria.

Pendeteksian konsumsi yang anomali ini dilakukan pada beberapa selang periode yang berurutan seperti pada Tabel 3.3 untuk mengetahui frekuensi kemunculan pelanggan terindikasi anomali. Apabila pelanggan tersebut mempunyai frekuensi kemunculan yang tinggi sebagai anomali, maka pelanggan tersebut kemungkinan besar adalah pelanggan yang anomali.

#### **4.1 Pre-Processing**

Sebelum melakukan analisa berdasarkan metode, *pre-processing* adalah tahap yang paling penting dimana data yang akan diolah disiapkan terlebih dahulu. *Pre-processing* yang dilakukan salah satunya adalah memeriksa *missing value* pada data pemakaian pelanggan PDAM Surya Sembada Kota Surabaya. Periode data tersebut dimulai pada Maret 2017 hingga Februari

2018. *Missing value* dapat terjadi jika terdapat pelanggan yang baru berlangganan atau berhenti berlanggan, sehingga tidak tercatat dibulan sesudahnya atau sebelumnya. Jumlah keseluruhan data disajikan pada Gambar 4.1 sebagai berikut.



**Gambar 4.1** Jumlah Pelanggan PDAM Surya Sembada Kota Surabaya Periode Maret 2017-Februari 2108

Gambar 4.1 menunjukkan bahwa terjadi peningkatan pelanggan dari waktu ke waktu. Supaya hasil analisa dapat maksimal, jumlah data masing-masing periode harus sama jumlahnya, sehingga dilakukan penyesuaian berdasarkan data yang paling sedikit, yaitu periode Maret 2017. Dengan kata lain, jika terdapat data yang tidak terdapat disalah satu periode penelitian, maka data tersebut dianggap *missing values* yang harus dihilangkan. Setelah dilakukan proses tersebut jumlah pelanggan yang terseisa pada penelitian ini adalah 546161 pelanggan.

Selanjutnya adalah melakukan pemeriksaan terhadap rumah kosong, karena penelitian ini hanya melakukan pemeriksaan anomali pada pelanggan aktif. Karena bangunan kosong yang berlangganan air juga cukup banyak di Kota Surabaya. Berikut adalah jumlah rumah kosong dengan rentang waktu tiga sampai dua belas bulan.

**Tabel 4.1** Jumlah Pelanggan Tak Aktif (Rumah Kosong)

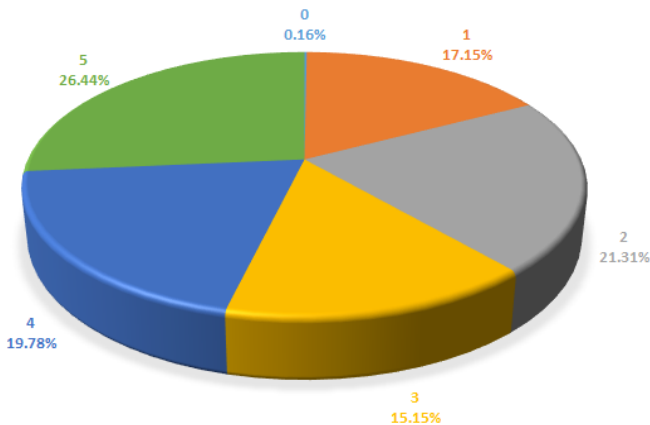
<b>Waktu</b>	<b>Jumlah Pelanggan</b>
3 Bulan	8060
4 Bulan	6316
5 Bulan	5168
6 Bulan	4586
7 Bulan	3877
8 Bulan	3540
9 Bulan	3324
10 Bulan	3563
11 Bulan	4603
12 Bulan	11650
Total	54687

Tabel 4.1 menunjukkan jumlah pelanggan yang tidak aktif atau rumah kosong dengan kriteria tidak terdapat catatan konsumsi air selama tiga hingga dua belas bulan berturut-turut. Sehingga terdapat 54687 rumah dengan konsumsi air 0 m<sup>3</sup>, dengan 11650 diantaranya adalah selama 12 bulan dan sisanya 3 hingga 11 bulan.

Selanjutnya adalah melakukan pemilihan golongan pelanggan, dimana golongan yang akan diteliti adalah golongan rumah tangga saja. Golongan Rumah tangga itu sendiri terbagi menjadi lima kelas berdasarkan tarif pembayaran. Pemebentukan tarif pelanggan tersebut berdasarkan kriteria tempat atau bangunan, daya listrik yang terpasang, nilai jual objek pajak, dan luas bangunan. Pada penelitian ini golongan tarif rumah tangga yang digunakan adalah golongan 2A II, 3A, 3C II, 4A, dan 4B II. Sehingga jumlah data tersisa yang digunakan sebanyak 454160 pelanggan. Kemudian langkah terakhir adalah mendapatkan nilai variabel-variabel yang ada pada Tabel 3.1 dan menyusun struktur data seperti pada Tabel 3.2, hasil akhir *pre-processing* dapat dilihat pada Lampiran 2.

#### 4.2 Karakteristik Data Pelanggan PDAM Surya Sembada Kota Surabaya

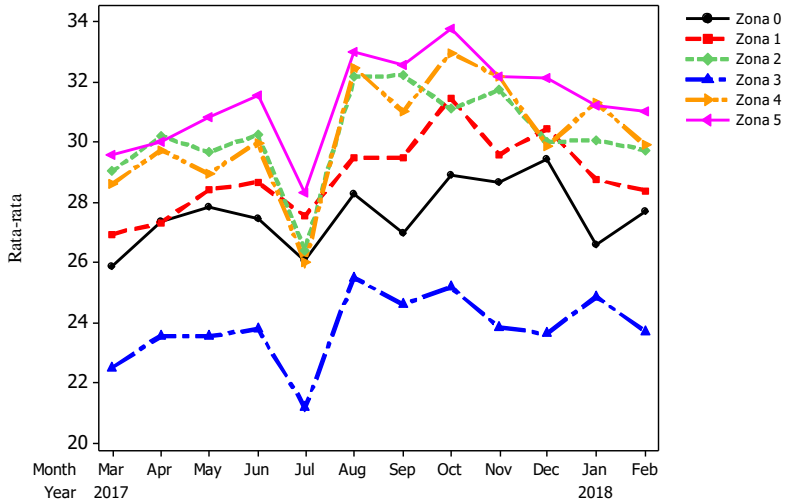
Pelanggan PDAM Surya Sembada Kota Surabaya terbagi kedalam 5 zona dan 148 sub-zona. Data yang digunakan hanya menggunakan golongan rumah tangga saja yang terbagi kedalam lima kode tarif pembayaran. Pada pembahasan karakteristik data hanya membahas jumlah pelanggan berdasarkan zona, dan rata-rata pemakaian. Berikut adalah jumlah pelanggan pada masing-masing zona.



**Gambar 4.2** Persentase Pelanggan PDAM Surya Sembada Kota Surabaya Berdasarkan Zona Periode April 2017 – Februari 2018

Gambar 4.2 menunjukkan jumlah pelanggan terbanyak adalah pada zona 6 dan yang paling sedikit adalah zona 5, dan yang paling sedikit pelanggan adalah zona 1. Dengan jumlah kurang lebih 120 ribu pelanggan. Selain ke-5 zona tersebut, terdapat zona untuk wilayah luar Surabaya, yaitu zona 0 dengan jumlah pelanggan sebanyak 710 pelanggan. Pembagian zona pelanggan PDAM didasarkan pada pompa-pompa/instalasi yang bekerja, bukan berdasarkan wilayah tempat tinggal seperti Surabaya Selatan, Utara, Barat, Timur, dan Pusat. PDAM Surya Sembada Kota Surabaya mempunyai 5 instalasi yakni Instalasi Ngagel I, Instalasi Ngagel II, Instalasi Ngagel III, Instalasi Kayoon dan Instalasi

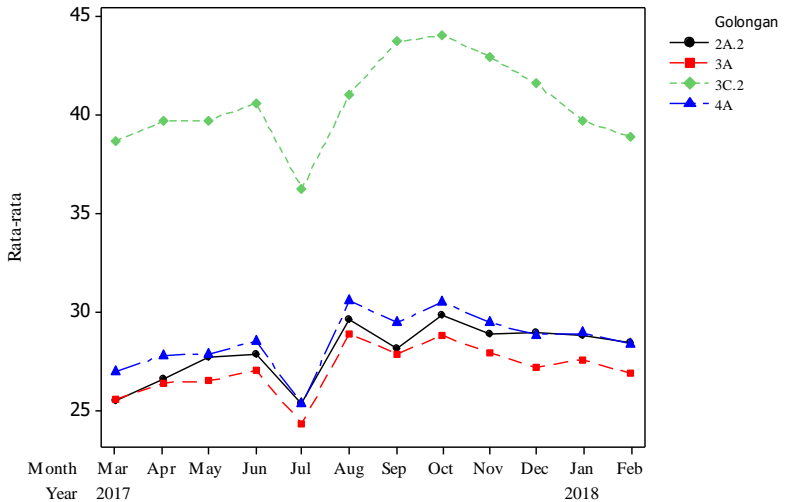
Karangpilang. Selanjutnya menghitung rata-rata berdasarkan Zona yang disajikan pada Gambar 4.3 berikut.



**Gambar 4.3** Rata-rata Pemakaian Pelanggan Berdasarkan Zona Periode April 2017 hingga Februari 2018

Gambar 4.3 menunjukkan bahwa pelanggan PDAM Surya Sembada Kota Surabaya memiliki pola pemakaian, yaitu selalu mengalami penurunan pada Bulan Juli baik di Zona 0 hingga Zona 5. Hal tersebut bias disebabkan adanya libur panjang atau hari-hari besar yang membuat penghuni tidak menempati rumah selama beberapa hari dan tidak melakukan konsumsi air. Gambar 4.3 juga menunjukkan bahwa konsumsi air terendah dimiliki oleh Zona 3 dengan rata-rata terendahnya pada Bulan Juli, dan tertinggi adalah Zona 5 dengan rata-rata tertingginya pada Bulan Oktober. Gambar 4.3 juga menunjukkan bahwa pola konsumsi pelanggan antar Zona diduga tidak memiliki perbedaan yang signifikan. Hal tersebut ditunjukkan bahwa rata-rata pemakaian berada dikisaran 26 hingga 30 m<sup>3</sup> di pada Zona 0, 1, 2, 4, dan 5.

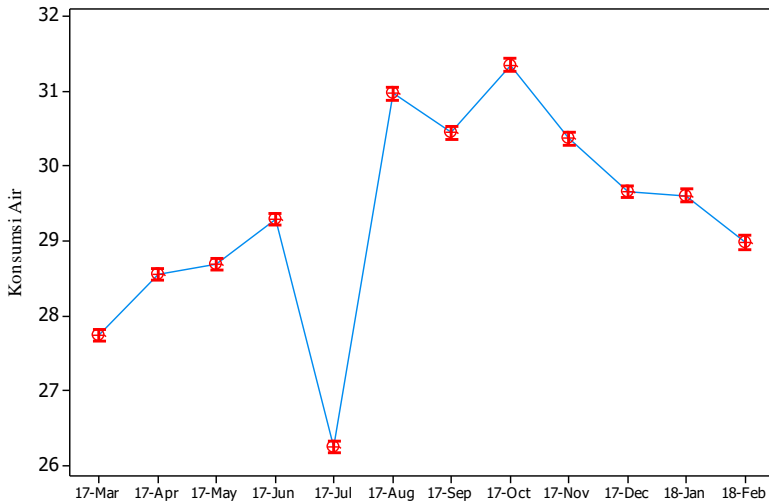
Selanjutnya adalah deskripsi pemakaian pelanggan berdasarkan golongan tarif. Pada kelas rumah tangga, terbagi menjadi 5 golongan tarif, yaitu 2A.II, 3A, 3C.II, 4A, dan 4B II. Berikut adalah pola pemakaian pelanggan berdasarkan golongan tariff yang disajikan pada Gambar 4.4.



**Gambar 4.4** Rata-rata Pemakaian Air PDAM Surya Sembada Kota Surabaya Berdasarkan Golongan Tarif.

Gambar 4.4 menunjukkan bahwa terdapat perbedaan rata-rata konsumsi pada golongan tarif 3C II. Golongan tarif tersebut mempunyai kriteria dengan bangunan yang berada dipinggir jalan protokol, di depan bangunan tersebut terdapat jalan dengan lebar termasuk saluran got lebih dari 15 meter, daya listrik yang terpasang lebih dari 4400 VA, nilai jual objek lebih dari 500 juta dan luas bangunan lebih dari 300m<sup>2</sup>. Dengan kata lain kategori golongan tarif tersebut adalah rumah tangga besar yang kemungkinan digunakan sebagai kantor atau industri. Sehingga membuktikan bahwa rata-rata pemakaian airnya berbeda dengan 3 golongan tarif lainnya.

Selanjutnya dapat dilihat pada Gambar 4.4 rata-rata pemakaian air selama 12 bulan



**Gambar 4.5** Rata-rata Pemakaian Air PDAM Surya Sembada Kota Surabaya Periode Maret 2017 – Februari 2018

Gambar 4.5 menunjukkan bahwa pola pemakaian air pelanggan mengalami kenaikan dari bulan Maret hingga Juni 2017, kemudian terjadi penurunan drastis pada Bulan Juli 2017. Setelah itu mengalami kenaikan pada Bulan Agustus 2017, dan terus mengalami penurunan hingga Februari 2018. Jika berdasarkan Gambar 4.5 terdapat dugaan anomali karena rata-rata pemakaian yang tidak begitu stabil. Secara keseluruhan rata-rata pemakaian air pelanggan berada di kisaran 30 m<sup>3</sup>.

#### 4.3 Pengelompokkan Menggunakan *Kohonen-SOM*

Pengelompokkan pada penelitian ini menggunakan 2 macam grid, dan dua macam topologi. Grid yang digunakan adalah grid 10x10 dan grid 20x20. Sehingga untuk grid 10x10 terdapat 100 grid dan untuk grid 20x20 terdapat 400 grid yang akan digunakan untuk mengelompokkan kurang lebih 450000 pelanggan. Data

yang digunakan dalam membentuk kelompok adalah data hasil *pre-processing* pada Lampiran 2. Topologi yang digunakan adalah topologi *gridtop* dan *hextop*. Sehingga terdapat 4 model yang terbentuk yaitu grid 10x10 dengan topologi *hextop* dan *gridtop* lalu grid 20x20 dengan topologi *hextop* dan *gridtop* pada masing-masing periode. Setelah membentuk model dengan masing-masing *grid* dan topologi, kemudian melakukan klastering berdasarkan nilai *code vector*. Nilai *code vector* adalah nilai yang digunakan untuk melakukan pengelompokan pada masing-masing *grid*, atau dengan kata lain ada nilai pusat masing-masing variabel di tiap *grid* untuk membentuk kelompok. Setelah didapatkan nilai *code vector* dilakukan pembentukan klaster 2 hingga 10 klaster berdasarkan *code vector* yang terbentuk. Penentuan jumlah klaster dipilih berdasarkan nilai *pseudo-F* dan ICD Rate. Nilai *pseudo-F* dan ICD Rate pada masing-masing model klaster terpilih dan masing-masing periode disajikan pada Tabel 4.2 berikut.

**Tabel 4.2** Nilai Pseudo-F dan ICD Rate Pada Klaster dan Model Terpilih

Periode	<i>Pseudo-F</i>	ICD Rate	Klaster Optimal
Mar 17 - Feb 18	38900.37	0.59	9
Apr 17 - Feb 18	48006.54	0.51	10
Mei 17 - Feb 18	27879.77	0.70	8
Jun 17 - Feb 18	47993.13	0.51	10
Jul 17 - Feb 18	47730.48	0.51	10
Agt 17 - Feb 18	40364.48	0.58	9
Sep 17 - Feb 18	45054.75	0.63	7
Okt 17 - Feb 18	53804.07	0.58	7
Nov 17 - Feb 18	65529.26	0.44	10
Des 17 - Feb 18	82011.44	0.38	10
Jan 18 - Feb 18	126545.8	0.34	8

Nilai *pseudo-F* yang dipilih adalah sebesar mungkin, karena nilai *pseudo-F* menunjukkan besarnya keragaman antar kelompok, sehingga semakin besar nilai *pseudo-F*, maka keragaman antar

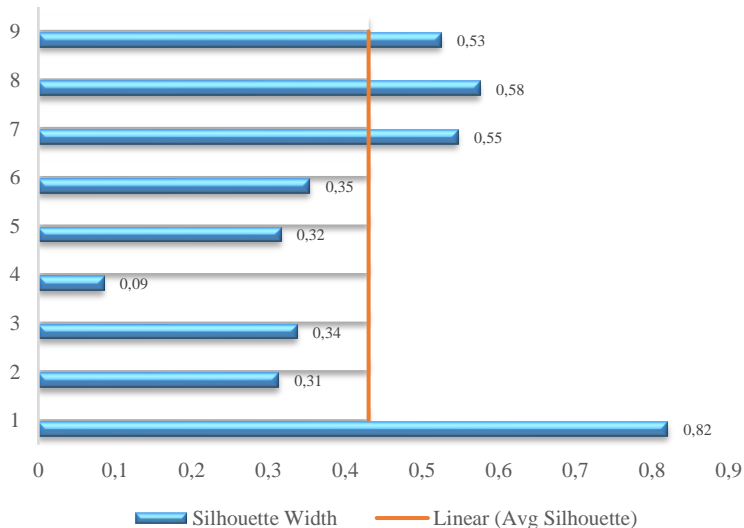


kelompok yang terbentuk juga semakin besar dan itu berarti baik. Sedangkan nilai *ICD Rate* yang diharapkan adalah sekecil mungkin, karena *ICD Rate* menggambarkan keragaman anggota dalam kelompok tersebut. Sehingga semakin kecil nilai *ICD Rate* maka semakin baik kelompok yang terbentuk. Tabel 4.2 merupakan hasil pembentukan kelompok menggunakan *vector code* yang dapat dilihat pada Lampiran 3. Tabel 4.2 menunjukkan bahwa klaster optimal yang terbentuk beragam. Umumnya nilai *pseudo-F* berkaitan dengan *ICD Rate*, semakin besar nilai *pseudo-F* biasanya juga semakin kecil nilai *ICD Rate*-nya. Seperti yang ditunjukkan pada Tabel 4.2 pada periode Mei 2017 hingga Februari 2018, nilai *pseudo-F* yang dihasilkan terkecil dibandingkan dengan periode lainnya, begitu juga dengan nilai *ICD Rate*-nya menunjukkan nilai yang paling besar dibandingkan periode lainnya. Langkah selanjutnya adalah memilih kelompok yang kurang baik menggunakan nilai *silhouette width* yang didapatkan dengan menggunakan persamaan 2.5 dari masing-masing klaster optimal yang telah terbentuk.

#### 4.4 Pemilihan Kelompok Anomali

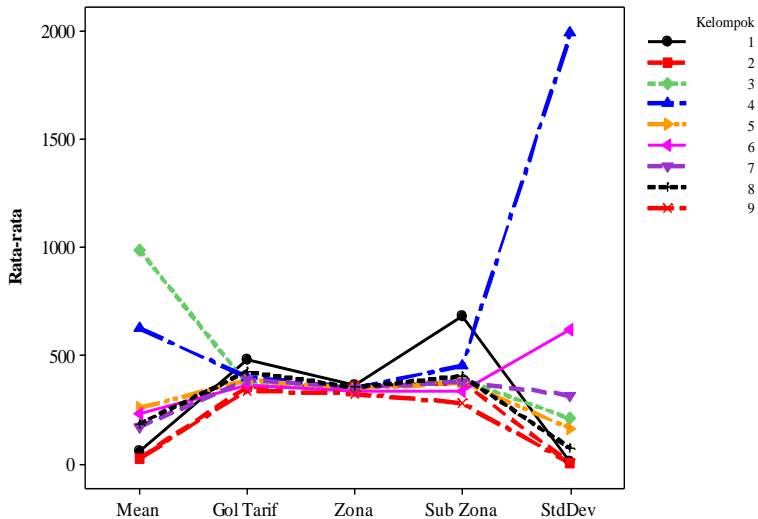
Pemilihan kelompok anomali menggunakan nilai *silhouette width*. Nilai *silhouette width* memiliki beberapa kriteria yang dijelaskan pada Tabel 2.1. Setelah didapatkan kelompok yang optimal dengan menggunakan algoritma *Kohonen SOM*, maka akan didapatkan hasil pengelompokan yang digunakan untuk menghitung nilai *silhouette*. Hasil kelompok yang digunakan dapat dilihat pada Lampiran 4. Nilai *silhouette* yang baik adalah mendekati 1. Jika nilai *silhouette* mendekati 1, dapat dikatakan keragaman anggota dalam kelompok yang terbentuk relatif kecil. Nilai *silhouette* yang didapatkan dapat dilihat pada Lampiran 5, setelah mendapatkan nilai *silhouette* pada masing-masing pengamatan di setiap periode, dihitung nilai *silhouette width* untuk masing-masing kelompok pada masing-masing periode. Batas yang digunakan untuk menentukan kelompok yang diduga anomali adalah rata-rata nilai *silhouette width* pada masing-masing kelompok yang terbentuk. Berikut adalah nilai *silhouette width*

pada masing-masing klaster yang terbentuk di masing-masing periode.



**Gambar 4.6** Nilai *Silhouette Width* 9 Klaster Periode Mar 17 – Feb 18

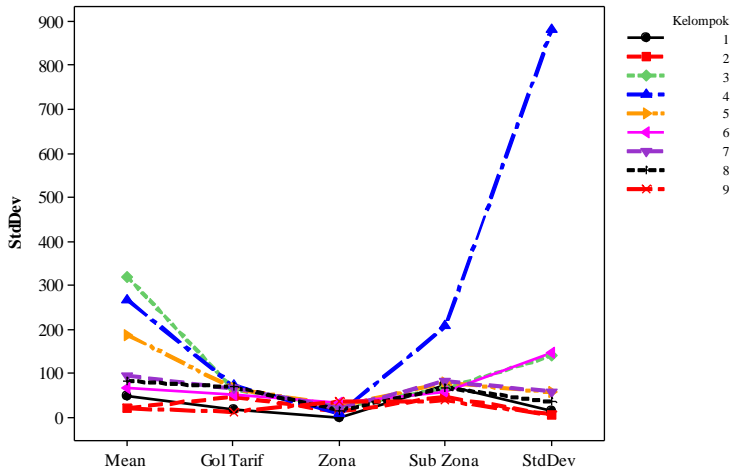
Gambar 4.6 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.43 yang berarti jika berdasarkan Tabel 2.1 kurang baik. Gambar 4.6 juga menunjukkan terdapat 5 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata. Empat dari lima kelompok tersebut masuk kedalam kriteria kelompok yang lemah karena nilai *silhouette width* tidak kurang dari 0.25, sedangkan satu kelompok yakni kelompok 4 mempunyai nilai *silhouette* sebesar 0.09 yang berarti tidak dapat disebut sebagai kelompok. Perhitungan nilai *silhouette* tersebut bertujuan untuk mendapatkan kelompok yang terduga anomali. Sehingga pada periode Maret 2017 hingga Februari 2018 diduga terdapat 5 kelompok yang terdapat anomali. Berikut adalah rata-rata masing-masing kelompok pada periode Maret 2017 hingga Februari 2018 yang disajikan pada Gambar 4.7.



**Gambar 4.7** Rata-rata Kelompok Periode Maret 2017 – Februari 2018

Gambar 4.7 menunjukkan nilai rata-rata pada masing-masing kelompok yang terbentuk. Sumbu  $x$  menjelaskan nilai rata-rata pada variabelnya. Pada variabel rata-rata pemakaian air (*Mean*), terdapat perbedaan rata-rata secara grafik pada masing-masing kelompok. Hal tersebut ditunjukkan tidak terdapat garis yang berhimpit pada masing-masing kelompok. Nilai rata-rata tertinggi pada variabel rata-rata pemakaian air (*Mean*) adalah kelompok 3 dan yang paling rendah adalah kelompok 9. Untuk variabel rata-rata pemakaian air berdasarkan golongan tarif (*Gol Tarif*) dan rata-rata pemakaian air berdasarkan zona (*Zona*) yang tertinggi adalah kelompok 1 dan yang paling rendah adalah kelompok 9. Untuk variabel rata-rata pemakaian air berdasarkan sub zona (*Sub Zona*) dan variabel deviasi standar pemakaian air (*StdDev*) terdapat perbedaan rata-rata secara grafik. Pada variabel deviasi standar pemakaian air (*StdDev*), nilai tertinggi adalah kelompok 4, sehingga dapat dikatakan bahwa kelompok mempunyai keragaman

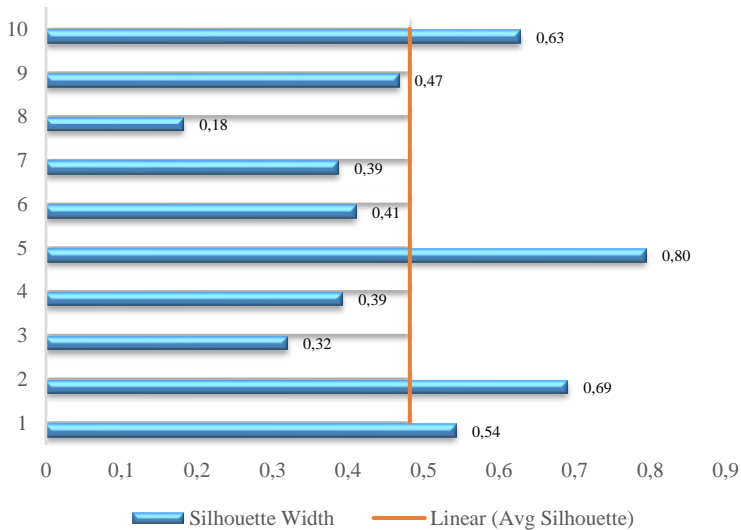
yang besar. Selanjutnya deviasi standar pada masing-masing variabel pada kelompok yang terbentuk periode Maret 2017 hingga Februari 2018 akan disajikan pada Gambar 4.8.



**Gambar 4.8** Deviasi Standar Kelompok Periode Maret 2017 – Februari 2018

Gambar 4.8 deviasi standar menunjukkan bahwa kelompok 4 mempunyai nilai deviasi standar yang paling tinggi pada variabel rata-rata pemakaian berdasarkan sub zona (*Sub Zona*) dan deviasi standar pemakaian air (*StdDev*). Kelompok 4 juga menghasilkan nilai *silhouette* yang kecil dibandingkan kelompok lainnya. Selain kelompok 4, kelompok 3 juga mempunyai deviasi standar yang tinggi pada variabel rata-rata pemakaian air (*Mean*), nilai *silhouette* yang didapatkan oleh kelompok 3 juga dibawah batas penentuan kelompok anomali. Sehingga berdasarkan Gambar 4.6 dan Gambar 4.8 kelompok yang mempunyai nilai *silhouette* yang kecil merupakan kelompok yang mempunyai keragaman yang tinggi pada variabel-variabel tertentu atau dapat dikatakan secara grafik kelompok yang terbentuk memiliki perbedaan pada rata-rata dan keragaman pada variabel-variabel tertentu. Kemudian akan

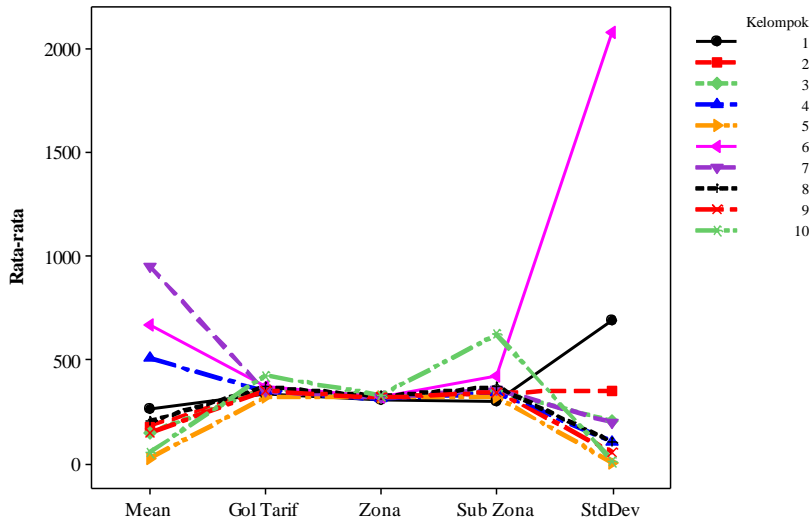
ditampilkan nilai *silhouette* dari periode April 2017 – Februari 2018 yang disajikan pada Gambar 4.8.



**Gambar 4.9** Nilai *Silhouette Width* 10 Klaster Periode Apr 17 – Feb 18

Gambar 4.9 menunjukkan 6 klaster dari 10 klaster memiliki nilai *silhouette width* dibawah rata-rata. Nilai *silhouette* rata-rata pada Gambar 4.9 adalah 0.48, yang berarti merupakan penentuan batas nilai *silhouette* suatu kelompok dikatakan kelompok anomali. Gambar 4.9 menunjukkan kelompok yang mempunyai nilai *silhouette* yang lebih dari batas adalah kelompok 1, kelompok 2, kelompok 5, dan kelompok 10. Sedangkan kelompok yang mempunyai nilai dibawah batas adalah kelompok 3, kelompok 4, kelompok 6, kelompok 7 dan kelompok 8. Kelompok 8 mempunyai nilai *silhouette* yang paling kecil dan termasuk kategori kedalam nilai *silhouette* kurang dari 0.25 atau tidak dapat dikatakan sebagai kelompok. Sedangkan 5 klaster lainnya masih tergolong kelompok yang lemah karena nilai *silhouette* tidak kurang dari 0.25. Nilai rata-rata pada kelompok yang terbentuk

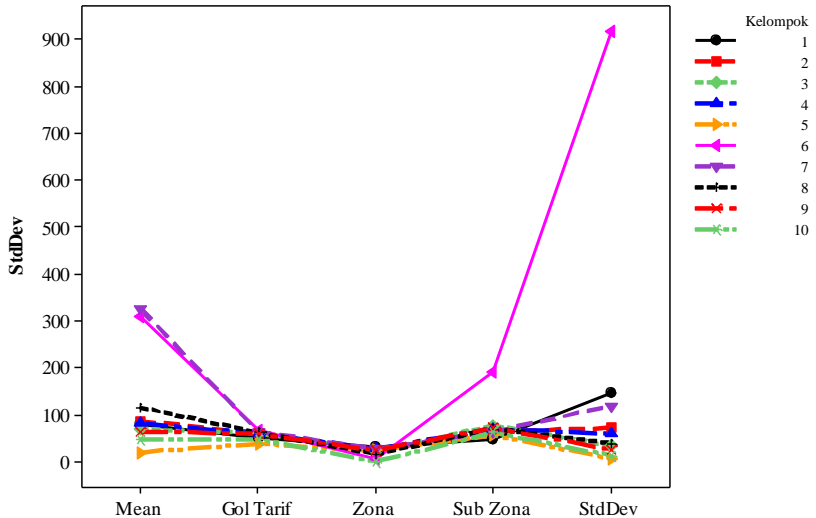
periode Mei 2017 hingga Februari 2018 disajikan pada Gambar 4.10.



**Gambar 4.10** Rata-rata Kelompok Periode April 2017 – Februari 2018

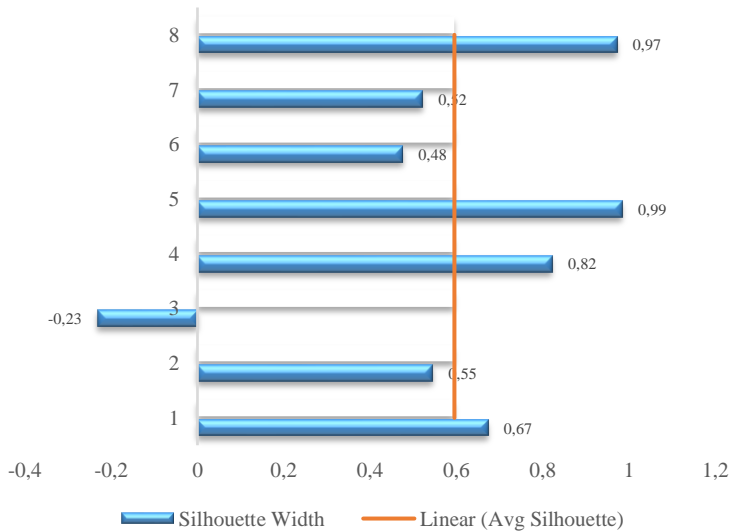
Gambar 4.10 menunjukkan nilai rata-rata masing-masing variabel pada periode Mei 2017 hingga Februari 2018. Rata-rata pada variabel rata-rata pemakaian air, rata-rata pemakaian air sub-zona dan deviasi standar pemakaian air terdapat perbedaan pada masing-masing kelompok yang terbentuk. Pada variabel rata-rata pemakaian air, kelompok 7 mempunyai rata-rata pemakaian air paling tinggi dibandingkan kelompok lainnya dan kelompok 5 mempunyai nilai rata-rata yang paling rendah pada variabel rata-rata pemakaian air hingga deviasi standar. Perbedaan rata-rata pada variabel rata-rata pemakaian berdasarkan tarif (*Gol Tarif*) dan variabel rata-rata pemakaian air berdasarkan zona (*Zona*) tidak begitu terlihat jelas seperti pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Secara kelompok yang terbentuk sudah memiliki perbedaan karakteristik berdasarkan variabel-variabelnya. Selanjutnya akan disajikan

deviasi standar pada kelompok yang terbentuk periode April 2017 hingga Februari 2018 pada Gambar 4.11.



**Gambar 4.11** Deviasi Standar Kelompok Periode April 2017 – Februari 2018

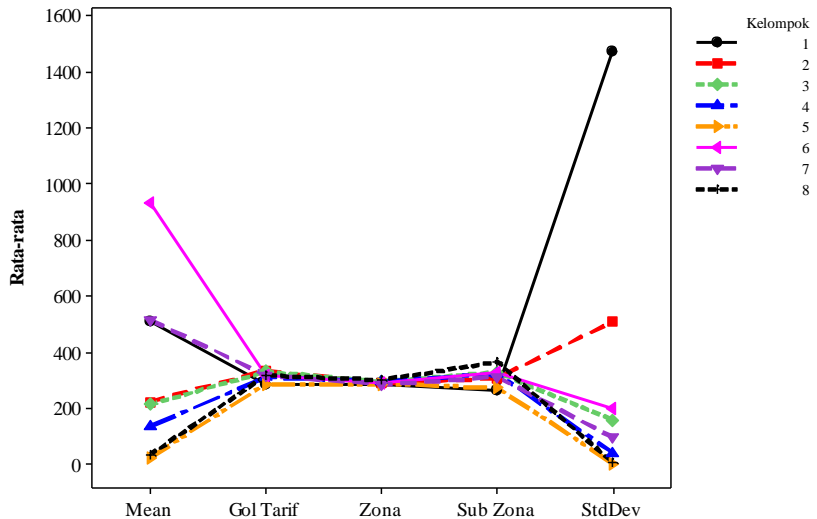
Gambar 4.11 menunjukkan terdapat perbedaan secara grafik deviasi standar pada masing-masing kelompok yang terbentuk pada periode April 2017 hingga Februari 2018. Jika dilihat pada deviasi standar, kelompok 7 mempunyai nilai deviasi standar yang paling tinggi dibandingkan kelompok lainnya pada variabel rata-rata pemakaian air (*Mean*) disusul dengan kelompok 6 dan kelompok 8. Kelompok 6 mempunyai deviasi standar yang paling tinggi pada 2 variabel, yaitu variabel sub zona dan variabel deviasi standar. Gambar 4.11 juga menunjukkan bahwa kelompok yang terbentuk pada periode April 2017 hingga Februari 2018 mempunyai perbedaan karakteristik deviasi standar walaupun hanya dilihat berdasarkan grafik saja. Berikutnya adalah nilai *silhouette width* periode Mei 2017 hingga Februari 2018 yang disajikan pada Gambar 4.12.



**Gambar 4.12** Nilai *Silhouette Width* 8 Klaster Periode Mei 17 – Feb 18

Gambar 4.12 menunjukkan nilai rata-rata *silhouette* yang terbentuk adalah 0.59 yang ditunjukkan dengan garis berwarna *orange*. Nilai rata-rata tersebut juga digunakan sebagai batas penentuan suatu kelompok dikatakan sebagai kelompok anomali. Gambar 4.12 menunjukkan terdapat 4 kelompok yang mempunyai nilai *silhouette* kurang dari batas yang telah ditentukan. Nilai *silhouette* yang paling kecil dihasilkan oleh kelompok 3 dan yang terbesar adalah kelompok 8. Jika dilihat berdasarkan nilai rata-rata *silhouette* sebesar 0.59, kelompok yang terbentuk sudah cukup baik dilihat dari keragaman antar kelompok dan juga keragaman antar anggota didalam kelompok tersebut. Berdasarkan Gambar 4.12 terdapat 4 kelompok yang diduga sebagai kelompok yang mengandung pola konsumsi tidak normal atau anomali. Selanjutnya akan disajikan grafik rata-rata pada kelompok yang terbentuk pada periode Mei 2017 hingga Februari 2018 pada Gambar 4.13.

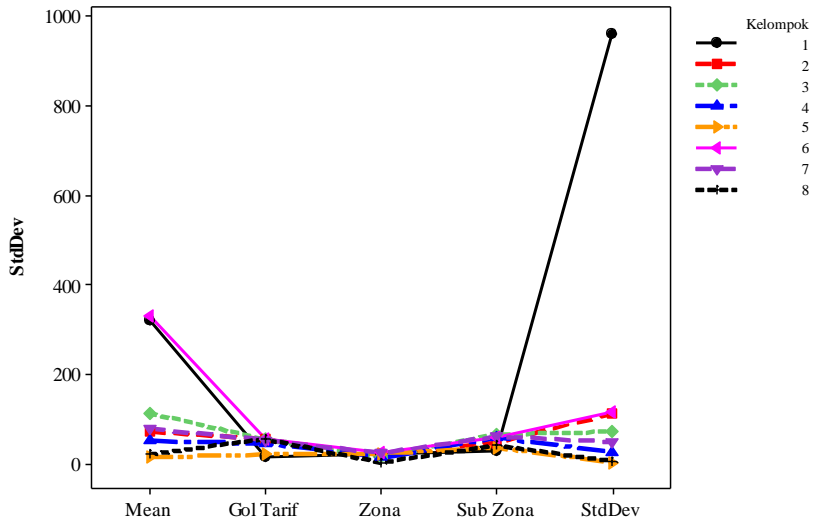




**Gambar 4.13** Rata-rata Kelompok Periode Mei 2017 – Februari 2018

Gambar 4.13 menunjukkan terdapat perbedaan secara grafik terhadap nilai rata-rata masing-masing kelompok. Pada variabel pemakaian rata-rata air (*Mean*), rata-rata pemakaian air berdasarkan sub zona (*Sub Zona*) dan deviasi standar pemakaian air (*StdDev*) perbedaan terlihat sangat jelas karena garis yang terbentuk saling terpisah. Kelompok 6 mempunyai nilai rata-rata tertinggi pada variabel rata-rata pemakaian air (*Mean*) disusul dengan kelompok 7 dan kelompok 1. Pada variabel rata-rata pemakaian air berdasarkan sub zona (*Sub Zona*) kelompok yang mempunyai nilai rata-rata paling tinggi adalah kelompok 8, dan pada variabel deviasi standar kelompok yang mempunyai nilai rata-rata tertinggi adalah kelompok 1. Berbeda dengan 3 variabel sebelumnya, variabel golongan tarif dan variabel zona tidak menunjukkan perbedaan yang sangat jelas secara grafik. Secara keseluruhan kelompok yang terbentuk pada periode Mei 2017 hingga Februari 2018 sudah memiliki perbedaan secara grafik. Selanjutnya akan ditampilkan karakteristik berdasarkan deviasi

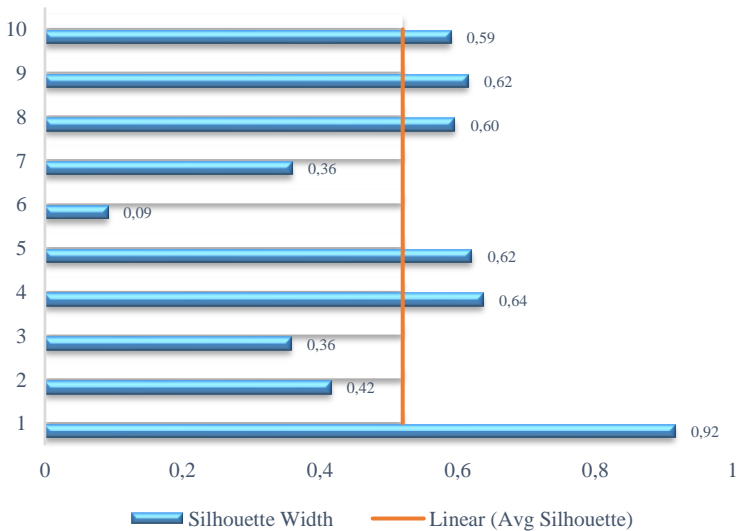
standar kelompok yang terbentuk periode Mei 2017 hingga Februari 2018 pada Gambar 4.14.



**Gambar 4.14** Deviasi Standar Kelompok Periode Mei 2017 – Februari 2018

Gambar 4.14 secara grafik menunjukkan adanya perbedaan deviasi standar atau keragaman antar kelompok. Perbedaan deviasi standar tersebut sangat jelas terlihat pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Pada variabel selain deviasi standar pemakaian air (*StdDev*), kelompok 6 dan 7 selalu mempunyai nilai keragaman yang tinggi. Pada variabel rata-rata pemakaian air (*Mean*) dan variabel rata-rata pemakaian air berdasarkan golongan tarif (*Gol Tarif*) keragaman yang paling tinggi adalah kelompok 6, variabel rata-rata pemakaian air berdasarkan zona (*Zona*) keragaman tertinggi terdapat pada kelompok 7 dan variabel pemakaian air berdasarkan sub zona (*Sub Zona*) keragaman tertinggi terdapat pada kelompok 3. Pada Gambar 4.12 juga ditunjukkan bahwa kelompok 3, 6, dan 7 memiliki nilai *silhouette* yang kecil dibawah batas yang telah

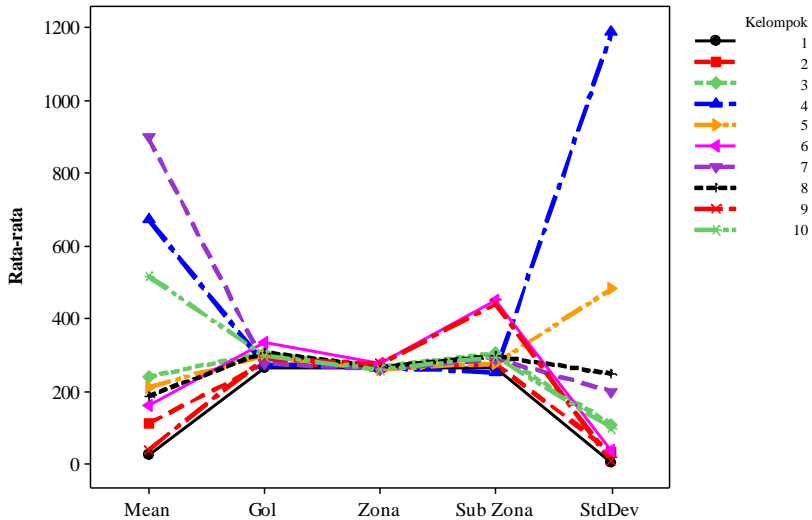
ditentukan. Secara grafik kelompok yang terbentuk telah memiliki perbedaan dalam hal deviasi standar. Berikutnya adalah nilai *silhouette* periode Juni 2017 hingga Februari 2018 yang disajikan pada Gambar 4.15.



**Gambar 4.15** Nilai *Silhouette Width* 10 Klaster Periode Juni 2017 – Februari 2018

Gambar 4.15 menunjukkan terdapat 4 klaster yang memiliki nilai *silhouette width* dibawah rata-rata. Nilai rata-rata *silhouette* pada Gambar 4.15 adalah sebesar 0.52, dengan begitu kelompok yang terbentuk atau klaster yang terbentuk termasuk kedalam kategori yang cukup baik. Garis *orange* pada Gambar 4.15 menunjukkan nilai rata-rata tersebut dan merupakan sebagai bagas penentuan kelompok yang diduga memiliki pola konsumsi tidak normal. Terdapat 4 kelompok yang diduga anomali dari 10 kelompok yang terbentuk, yaitu kelompok 2, kelompok 3, kelompok 6, dan kelompok 7. Nilai *silhouette* yang paling kecil terdapat pada kelompok 6 yaitu sebesar 0.09 dan termasuk kedalam kategori yang sangat buruk untuk dikatakan sebagai kelompok.

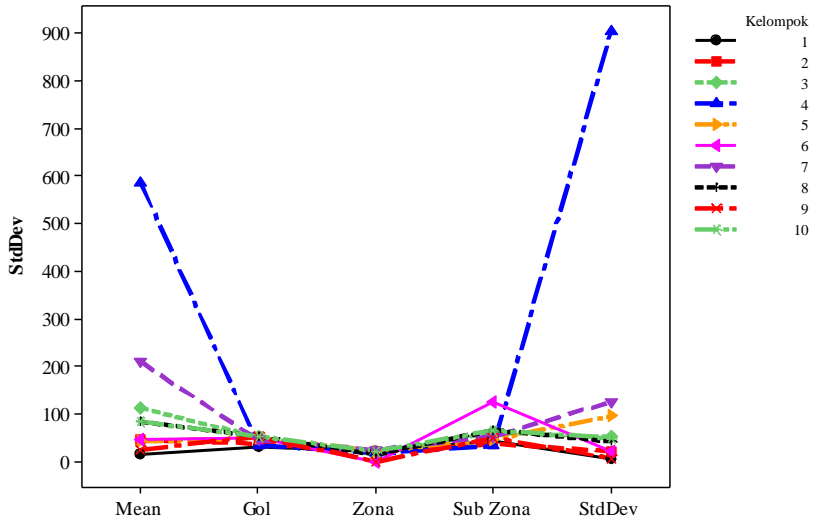
Sedangkan 3 klaster lainnya masih dapat dikatakan sebagai kelompok yang lemah karena tidak kurang dari 0.25. Selanjutnya disajikan karakteristik berupa rata-rata pada Gambar 4.16 untuk masing-masing kelompok yang terbentuk.



**Gambar 4.16** Rata-rata Kelompok Periode Juni 2017 – Februari 2018

Gambar 4.16 menunjukkan perbedaan rata-rata yang paling jelas terlihat pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Kelompok 7 memiliki nilai rata-rata yang paling tinggi pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*), kelompok 6 pada variabel rata-rata pemakaian air berdasarkan golongan tarif, serta variabel rata-rata pemakaian air berdasarkan zona dan sub zona. Secara grafik, kelompok yang terbentuk dapat dikatakan memiliki karakteristik *mean* yang berbeda pada masing-masing variabel, hanya saja perbedaan yang paling jelas terlihat pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Selanjutnya Gambar 4.17 menyajikan

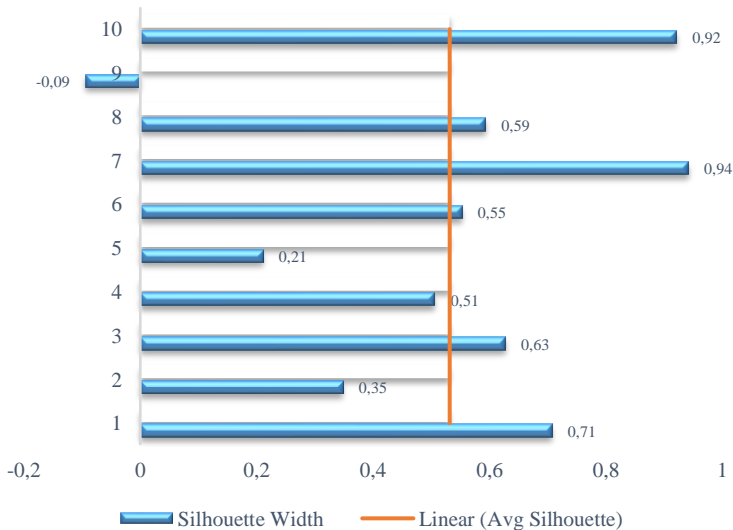
karakteristik deviasi standar masing-masing kelompok periode Juni 2017 hingga Februari 2018.



**Gambar 4.17** Deviasi Standar Kelompok Periode Juni 2017 – Februari 2018

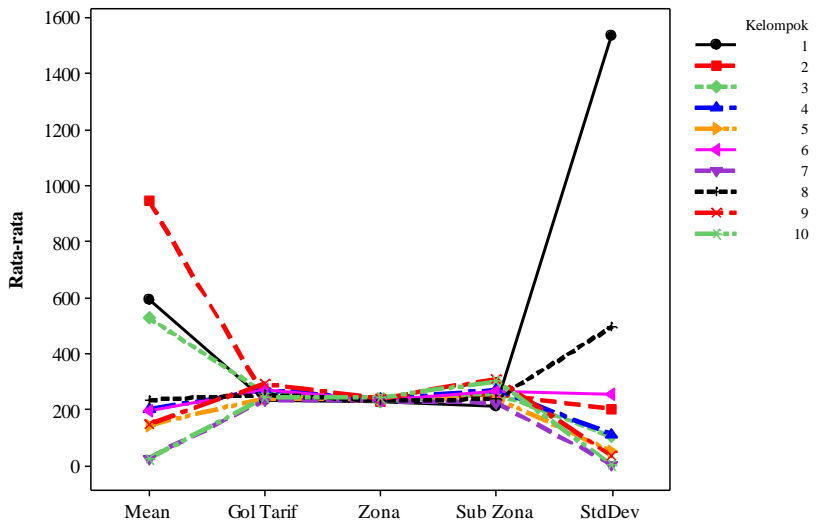
Gambar 4.17 menunjukkan deviasi standar pada masing-masing kelompok periode Juni 2017 hingga Februari 2018 telah berbeda secara grafik. Perbedaan yang paling jelas terdapat pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Kelompok 4 mempunyai deviasi standar yang paling tinggi di dua variabel, yakni variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air (*StdDev*). Namun nilai *silhouette* yang dihasilkan kelompok 4 juga cukup baik, sehingga besarnya deviasi standar kelompok 4 pada rata-rata pemakaian air (*Mean*) dikarenakan anggota dalam kelompok tersebut memiliki rata-rata pemakaian air yang besar dibandingkan kelompok lainnya. Serta besarnya deviasi standar pada kelompok 4 hanya terjadi di dua variabel saja, oleh karena itu nilai *silhouette* yang dihasilkan masih cukup baik. Selain kelompok

4, kelompok 7 mempunyai deviasi standar yang tinggi di beberapa variabel, yaitu variabel rata-rata pemakaian air (*Mean*), variabel rata-rata pemakaian air berdasarkan golongan (*Gol Tarif*), variabel rata-rata pemakaian air berdasarkan zona (*Zona*) dan variabel deviasi standar pemakaian air. Berbeda dengan kelompok 4 yang hanya mempunyai nilai keragaman tinggi dikarenakan anggota dalam kelompoknya mempunyai nilai rata-rata pemakaian air yang memang tinggi dan hanya pada dua variabel saja. Kelompok 7 menggambarkan bahwa anggota didalam kelompok tersebut memang memiliki keragaman yang besar, sehingga nilai *silhouette* yang dihasilkan dari kelompok 7 kecil dan tidak lebih dari batas yang ada. Gambar 4.17 juga menunjukkan bahwa pembentukan kelompok kedalam 10 kelompok pada periode Juni 2017 hingga Februari 2018 telah memiliki perbedaan deviasi standar secara grafik. Selanjutnya disajikan Gambar 4.18 terkait nilai *silhouette* periode Juli 2017 hingga Februari 2018..



**Gambar 4.18** Nilai *Silhouette Width* 10 Klaster Periode Juli 17 – Feb 18

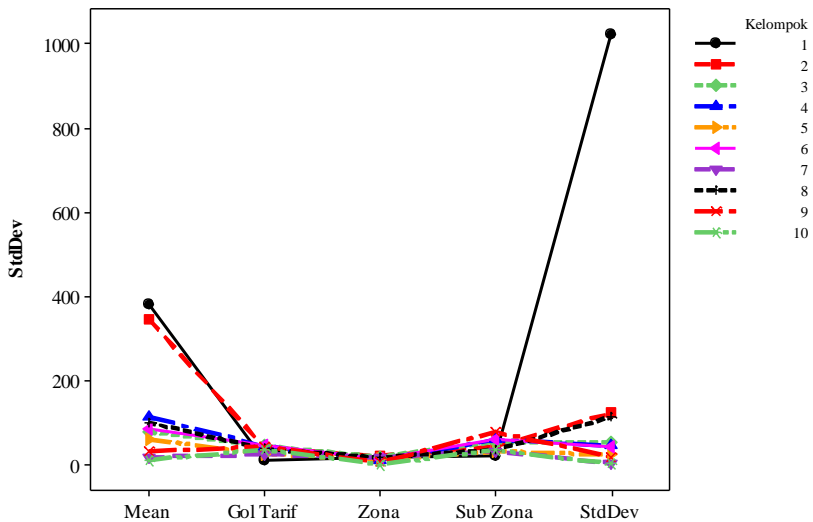
Gambar 4.18 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.53 yang berarti jika berdasarkan Tabel 2.1 cukup baik. Garis berwarna *orange* menunjukkan nilai rata-rata tersebut yang digunakan sebagai batas penentuan suatu kelompok dikatakan sebagai anomali. Gambar 4.18 juga menunjukkan terdapat 4 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata. 1 dari 4 kelompok tersebut masuk kedalam kriteria kelompok yang cukup baik karena nilai *silhouette width* tidak kurang dari 0.5, sedangkan dua kelompok yakni kelompok 9 dan kelompok 5 mempunyai nilai *silhouette* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Selanjutnya disajikan Gambar 4.19 yang menunjukkan karakteristik rata-rata masing-masing kelompok pada periode Juli 2017 hingga Februari 2018.



**Gambar 4.19** Rata-rata Kelompok Periode Juli 2017 – Februari 2018

Secara graifik Gambar 4.19, kelompok yang terbentuk telah mempunyai rata-rata yang berbeda pada masing-masing variabel.

Perbedaan rata-rata yang terlihat jelas ditunjukkan pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Gambar 4.19 juga menunjukkan kelompok 2 mempunyai rata-rata yang paling tinggi pada variabel rata-rata pemakaian air (*Mean*) dan disusul oleh kelompok 1. Seperti halnya periode-periode sebelumnya, perbedaan rata-rata pada masing-masing kelompok hanya terlihat jelas pada variabel pemakaian rata-rata air dan deviasi standar pemakaian air. Selanjutnya disajikan Gambar 4.20 yang menampilkan deviasi standar kelompok periode Juli 2017 hingga Februari 2018.

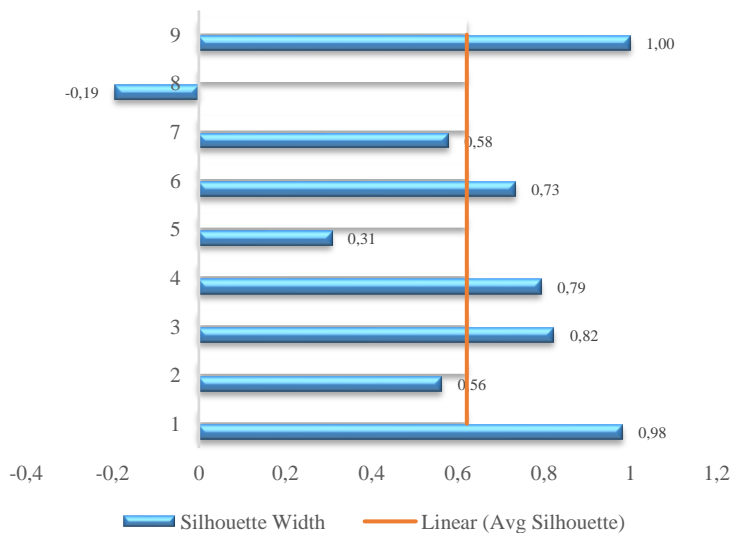


**Gambar 4.20** Deviasi Standar Kelompok Periode Juli 2017 – Februari 2018

Seperti halnya pada Gambar 4.19, Gambar 4.20 juga menunjukkan kelompok 2 mempunyai deviasi standar yang tinggi pada variabel rata-rata pemakaian air (*Mean*). Gambar 4.20 juga menunjukkan perbedaan deviasi standar terlihat jelas pada variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air. Kelompok 2 juga mempunyai nilai deviasi standar



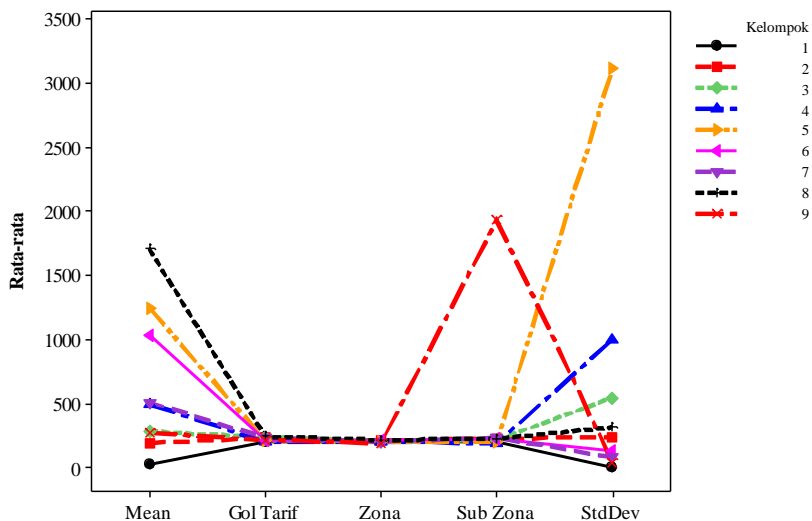
tertinggi pada variabel rata-rata pemakaian air berdasarkan golongan tarif (*Gol Tarif*). Selain kelompok 2, kelompok 1 mempunyai deviasi standar yang tinggi pada dua variabel, yaitu variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air. Hal tersebut serupa pada kelompok 4 periode Juni 2017 hingga Februari 2018 yang disebabkan oleh anggota kelompok didalamnya mempunyai rata-rata pemakaian air yang tinggi dibandingkan kelompok lainnya. Secara grafik kelompok yang terbentuk pada periode Agustus 2017 hingga Februari 2018 telah mempunyai deviasi standar yang berbeda. Berikutnya akan ditampilkan nilai *silhouette* dari periode Agustus 2017 – Februari 2018 yang disajikan pada Gambar 4.21



**Gambar 4.21** Nilai *Silhouette Width* 9 Klaster Periode Agustus 17 – Februari 18

Gambar 4.21 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.62 yang ditunjukkan oleh garis berwarna *orange* dan jika berdasarkan Tabel 2.1 cukup baik. Rata-rata tersebut digunakan sebagai batas penentuan suatu kelompok dikatakan sebagai

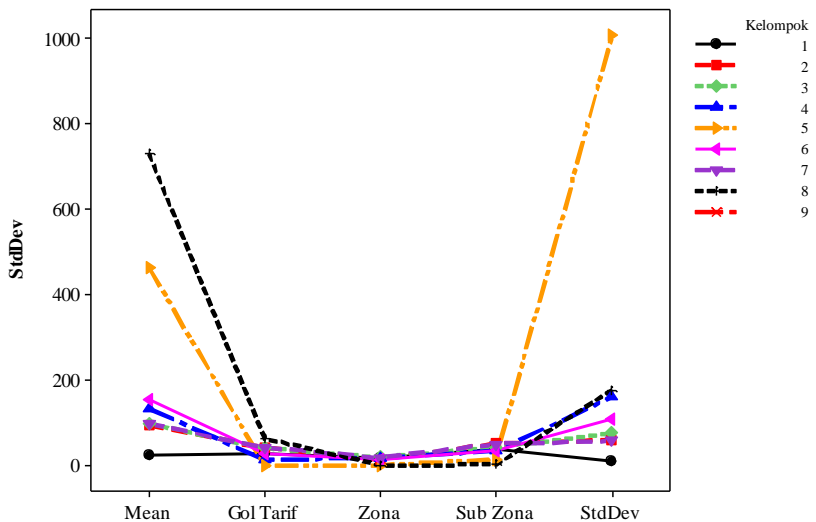
anomali atau tidak. Gambar 4.10 juga menunjukkan terdapat 4 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata 2 dari 4 kelompok tersebut masuk kedalam kriteria kelompok yang cukup baik karena nilai *silhouette width* tidak kurang dari 0.5, sedangkan 1 kelompok yakni kelompok 8 mempunyai nilai *silhouette* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Pada periode ini terdapat klaster yang hanya mempunyai satu anggota saja, yaitu klaster 9. Karakteristik rata-rata masing-masing kelompok dapat dilihat pada Gambar 4.22 berikut.



**Gambar 4.22** Rata-rata Kelompok Periode Agustus 2017 – Februari 2018

Gambar 4.22 mempunyai pola yang sedikit berbeda dari periode-periode sebelumnya. Gambar 4.22 menunjukkan perbedaan rata-rata secara grafik pada variabel rata-rata pemakaian (*Mean*) dan deviasi standar pemakaian air. Lebih dari itu, pada variabel pemakaian air berdasarkan sub zona terdapat kelompok dengan nilai rata-rata yang tinggi dikarenakan hanya terdapat 1

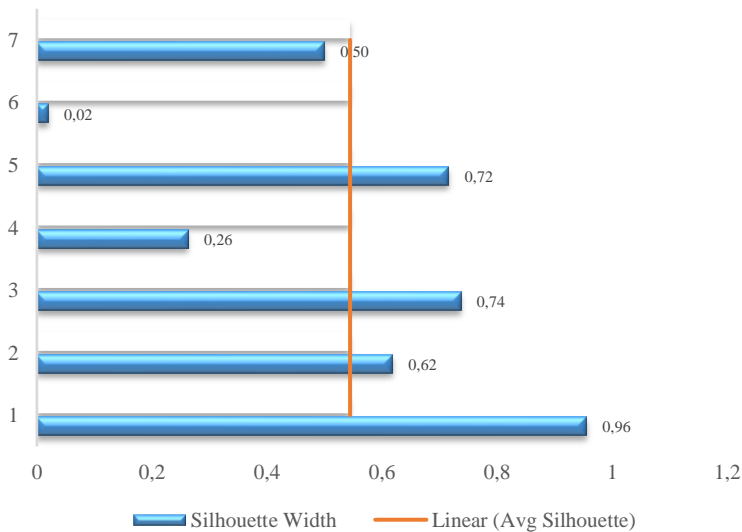
anggota saja didalam kelompok tersebut, yaitu kelompok 9. Kelompok 8 mempunyai rata-rata tertinggi pada variabel rata-rata pemakaian air (*Mean*) dan kelompok 5 mempunyai rata-rata tertinggi pada deviasi standar pemakaian air. Selanjutnya dapat dilihat karakteristik deviasi standar kelompok yang terbentuk pada periode Agustus 2017 hingga Februari 2018 yang disajikan pada Gambar 4.23 berikut.



**Gambar 4.23** Deviasi Standar Kelompok Periode Agustus 2017 Hingga Februari 2018

Gambar 4.23 menunjukkan deviasi standar kelompok periode Agustus 2017 hingga Februari 2018 telah berbeda secara grafik. Terutama pada variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air. Kelompok 8 mempunyai deviasi standar yang paling tinggi pada variabel rata-rata pemakaian air (*Mean*) kemudian disusul oleh kelompok 5. Pada variabel pemakaian air berdasarkan golongan tarif (*Gol Tarif*) kelompok 8 juga mempunyai deviasi standar yang paling tinggi. Kemudian kelompok 5 mempunyai deviasi standar paling tinggi

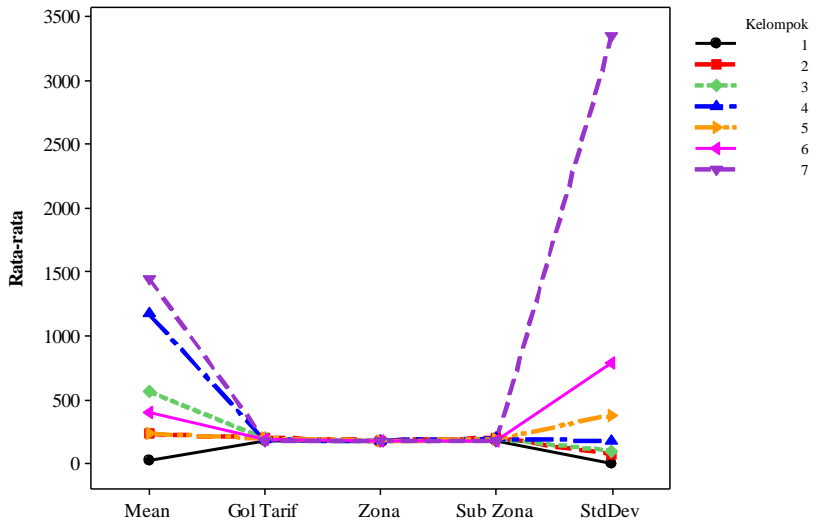
pada variabel deviasi standar pemakaian air. Pada Gambar 4.21 kedua kelompok tersebut mempunyai nilai *silhouette* yang kurang dari batas rata-rata. Secara grafik kelompok yang terbentuk pada periode Agustus 2017 hingga Februari 2018 telah berbeda dalam hal deviasi standar atau keragaman. Berikutnya akan ditampilkan nilai *silhouette* dari periode September 2017 – Februari 2018 yang disajikan pada Gambar 4.24.



**Gambar 4.24** Nilai *Silhouette Width* 7 Klaster Periode Sep 17 – Feb 18

Gambar 4.24 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.54 yang berarti jika berdasarkan Tabel 2.1 cukup baik. Gambar 4.24 juga menunjukkan terdapat 3 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata. 1 dari 3 kelompok tersebut masuk kedalam kriteria kelompok yang cukup baik karena nilai *silhouette width* tidak kurang dari 0.5, sedangkan 1 kelompok yakni kelompok 6 mempunyai nilai *silhouette* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Kemudian kelompok 4 dengan nilai *silhouette* sebesar

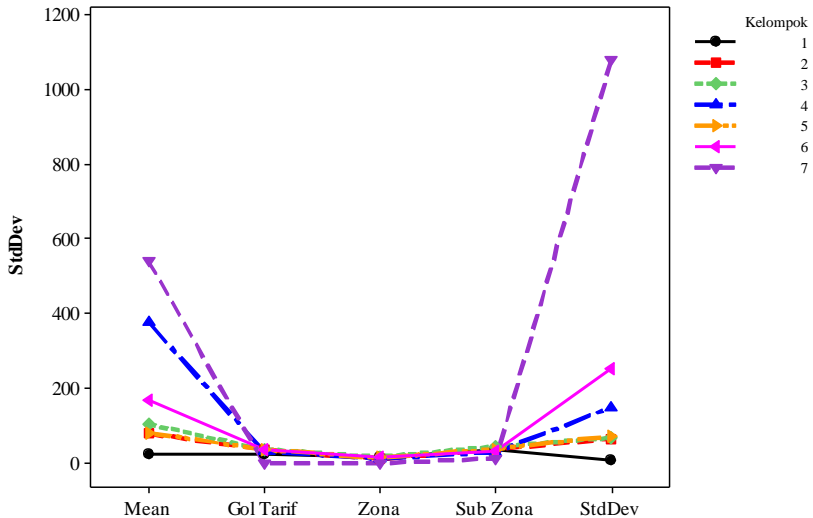
0.26 masuk kedalam kategori kelompok yang lemah. Selanjutnya disajikan karakteristik rata-rata pada masing-masing kelompok yang terbentuk periode September 2017 hingga Februari 2018 pada Gambar 4.25 berikut.



**Gambar 4.25** Rata-rata Kelompok Periode September 2017 – Februari 2018

Gambar 4.25 menunjukkan bahwa terdapat perbedaan secara grafik pada nilai rata-rata tiap kelompok pada periode September 2017 hingga Februari 2018. Perbedaan yang paling jelas terlihat pada variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air (*StdDev*). Kelompok 7 mempunyai nilai rata-rata yang paling besar dibandingkan kelompok lainnya pada variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air (*StdDev*). Kemudian disusul oleh kelompok 4 pada variabel rata-rata pemakaian air (*Mean*) dan kelompok 6 pada variabel deviasi standar pemakaian air (*StdDev*). Secara grafik, kelompok yang terbentuk pada periode September 2017 hingga Februari 2018 telah berbeda dalam hal rata-rata walaupun

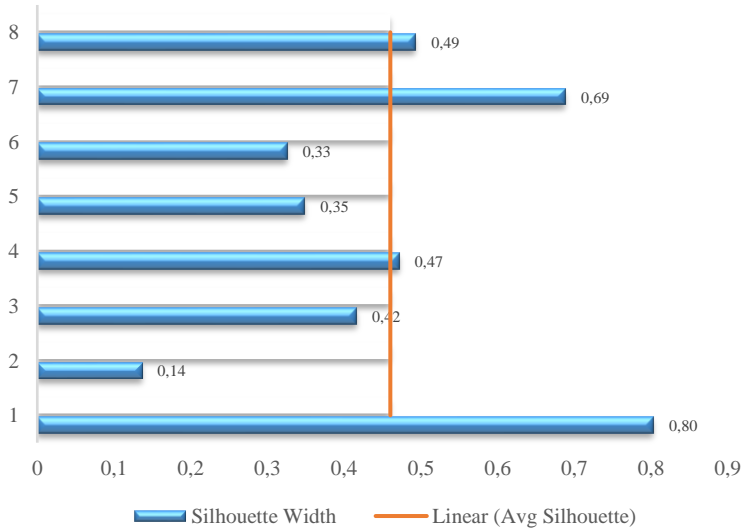
hanyak dibuktikan secara grafik saja. Selanjutnya adalah karakteristik deviasi standar kelompok periode September 2017 hingga Februari 2018 yang disajikan pada Gambar 4.26 berikut.



**Gambar 4.26** Deviasi Standar Kelompok Periode September 2017 – Februari 2018

Gambar 4.26 secara grafik menunjukkan adanya perbedaan deviasi standar atau keragaman antar kelompok. Perbedaan deviasi standar tersebut sangat jelas terlihat pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*), kelompok 7 selalu mempunyai nilai keragaman yang tinggi. Pada variabel rata-rata pemakaian air berdasarkan golongan tarif (*Gol Tarif*), dan variabel rata-rata pemakaian air berdasarkan zona (*Zona*) keragaman yang paling tinggi adalah kelompok 6. Pada Gambar 4.24 juga ditunjukkan bahwa kelompok 3, 6, dan 7 memiliki nilai *silhouette* yang kecil dibawah batas yang telah ditentukan. Secara grafik kelompok yang terbentuk telah memiliki perbedaan dalam hal deviasi standar.

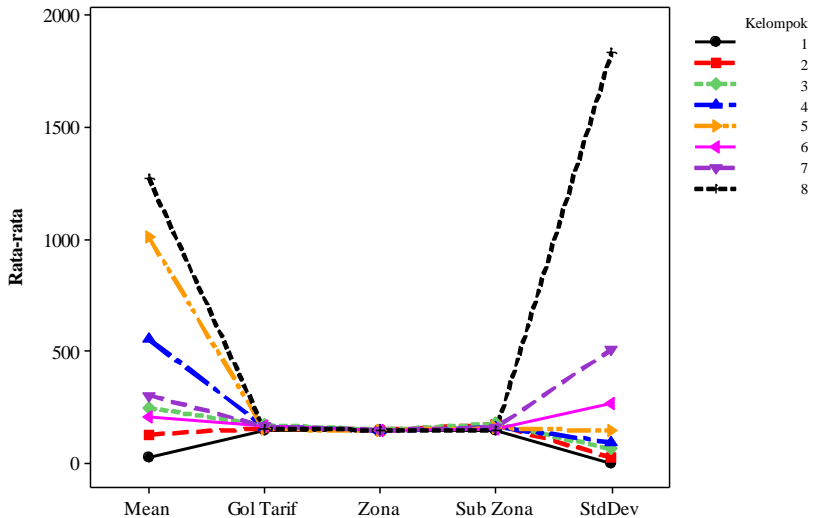
Berikutnya akan ditampilkan nilai *silhouette* dari periode Oktober 2017 – Februari 2018 yang disajikan pada Gambar 4.27.



**Gambar 4.27** Nilai *Silhouette Width* 8 Klaster Periode Okt 17 – Feb 18

Gambar 4.27 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.46 yang berarti jika berdasarkan Tabel 2.1 pengelompokkan tersebut membentuk kelompok dengan mayoritas kategori kelompok yang lemah. Nilai rata-rata tersebut ditunjukkan oleh garis berwarna *orange* pada Gambar 4.27 dan digunakan sebagai batas pemiliah suatu kelompok dikatakan kelompok anomali. Gambar 4.27 juga menunjukkan terdapat 4 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata atau batas, 3 dari 4 kelompok tersebut masuk kedalam kriteria kelompok yang yang lemah karena nilai *silhouette width* kurang dari 0.5, sedangkan 1 kelompok yakni kelompok 2 mempunyai nilai *silhouette* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Selanjutnya disajikan karakteristik berupa rata-rata dan deviasi standar pada masing-masing kelompok, Gambar

4.28 menunjukkan rata-rata pada masing-masing kelompok yang terbentuk pada periode Oktober 2017 hingga Februari 2018.

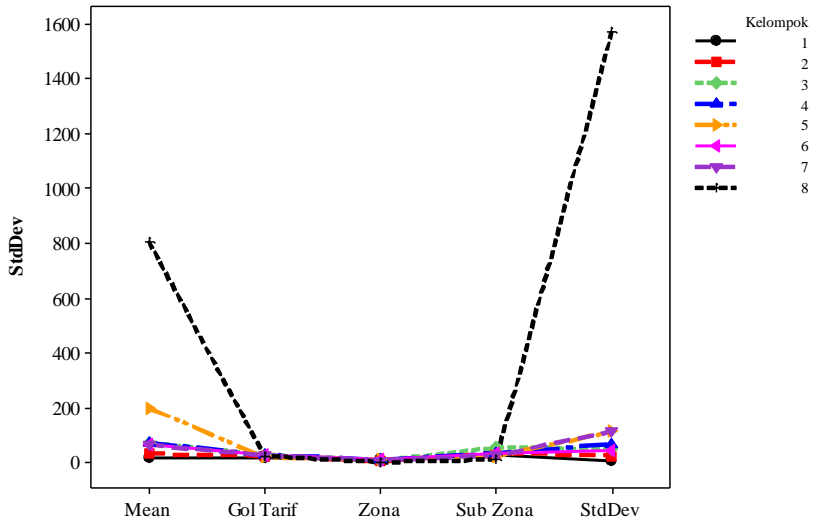


**Gambar 4.28** Rata-rata Kelompok Periode Oktober 2017 – Februari 2018

Gambar 4.28 menunjukkan terdapat perbedaan secara grafik terhadap nilai rata-rata masing-masing kelompok. Pada variabel pemakaian rata-rata air (*Mean*), dan deviasi standar pemakaian air (*StdDev*) perbedaan terlihat sangat jelas karena garis yang terbentuk saling terpisah. Kelompok 8 mempunyai nilai rata-rata tertinggi pada variabel rata-rata pemakaian air (*Mean*) disusul dengan kelompok 5 dan kelompok 4. Pada variabel rata-rata pemakaian air berdasarkan sub zona (*Sub Zona*) kelompok yang mempunyai nilai rata-rata paling tinggi adalah kelompok 3, dan pada variabel deviasi standar kelompok yang mempunyai nilai rata-rata tertinggi adalah kelompok 8. Berbeda dengan 2 variabel sebelumnya, variabel golongan tarif, variabel zona dan variabel sub zona tidak menunjukkan perbedaan yang sangat jelas secara grafik. Secara keseluruhan kelompok yang terbentuk pada periode



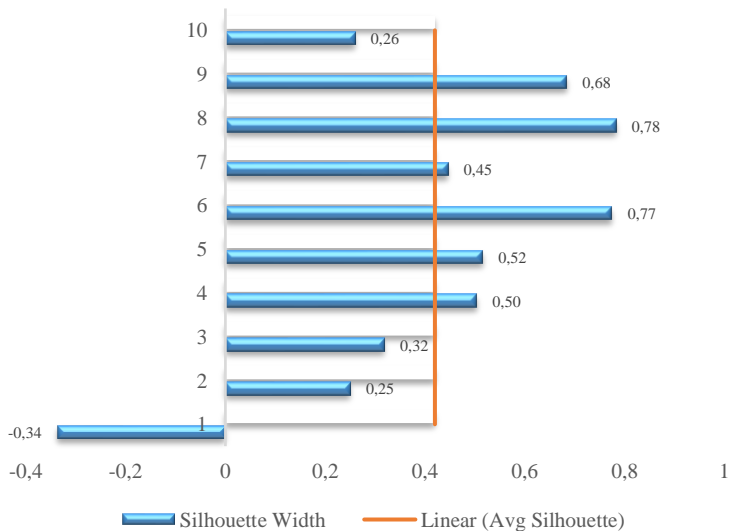
Oktober 2017 hingga Februari 2018 sudah memiliki perbedaan secara grafik. Selanjutnya akan ditampilkan karakteristik berdasarkan deviasi standar kelompok yang terbentuk periode Oktober 2017 hingga Februari 2018 pada Gambar 4.14



**Gambar 4.29** Deviasi Standar Kelompok Periode Oktober 2017 – Februari 2018

Gambar 4.29 menunjukkan terdapat perbedaan secara grafik deviasi standar pada masing-masing kelompok yang terbentuk pada periode Oktober 2017 hingga Februari 2018. Jika dilihat pada deviasi standar, kelompok 8 mempunyai nilai deviasi standar yang paling tinggi dibandingkan kelompok lainnya pada variabel rata-rata pemakaian air (*Mean*) disusul dengan kelompok 5 dan kelompok 6. Kelompok 8 mempunyai deviasi standar yang paling tinggi pada 2 variabel, yaitu variabel rata-rata pemakaian air dan variabel deviasi standar. Gambar 4.29 juga menunjukkan bahwa kelompok yang terbentuk pada periode Oktober 2017 hingga Februari 2018 mempunyai perbedaan karakteristik deviasi standar walaupun hanya dilihat berdasarkan grafik saja. Berikutnya adalah

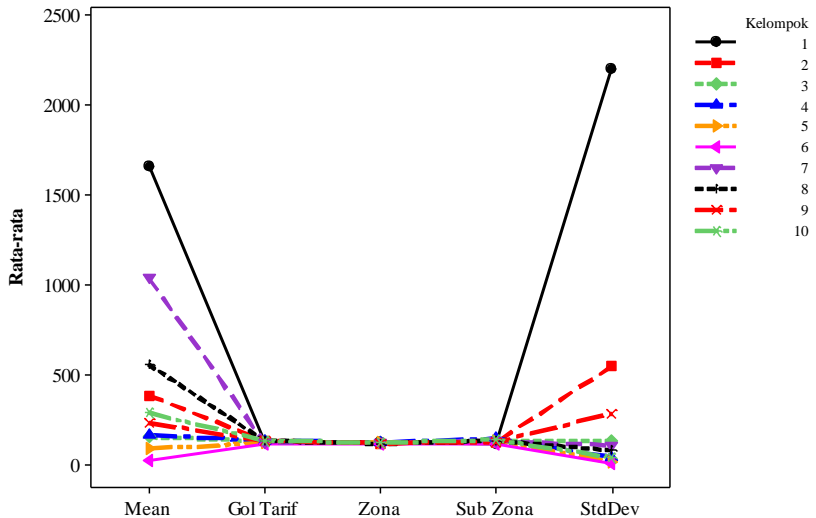
nilai *silhouette width* periode Mei 2017 hingga Februari 2018 yang disajikan pada Gambar 4.12. Berikutnya akan ditampilkan nilai *silhouette* dari periode November 2017 – Februari 2018 yang disajikan pada Gambar 4.30.



**Gambar 4.30** Nilai *Silhouette Width* 10 Klaster Periode Nov 17–Feb 18

Gambar 4.30 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.42 yang berarti jika berdasarkan Tabel 2.1 mayoritas kelompok yang terbentuk masuk kedalam kategori kelompok yang lemah. Garis berwarna *orange* menunjukkan batas yang digunakan untuk menentukan suatu kelompok dikatakan atau diduga sebagai kelompok anomali. Gambar 4.30 juga menunjukkan terdapat 4 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata. 3 dari 4 kelompok tersebut masuk kedalam kriteria kelompok yang lemah karena nilai *silhouette width* kurang dari 0.5, sedangkan 1 kelompok yakni kelompok 1 mempunyai nilai *silhouette* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Selanjutnya dapat dilihat rata-rata

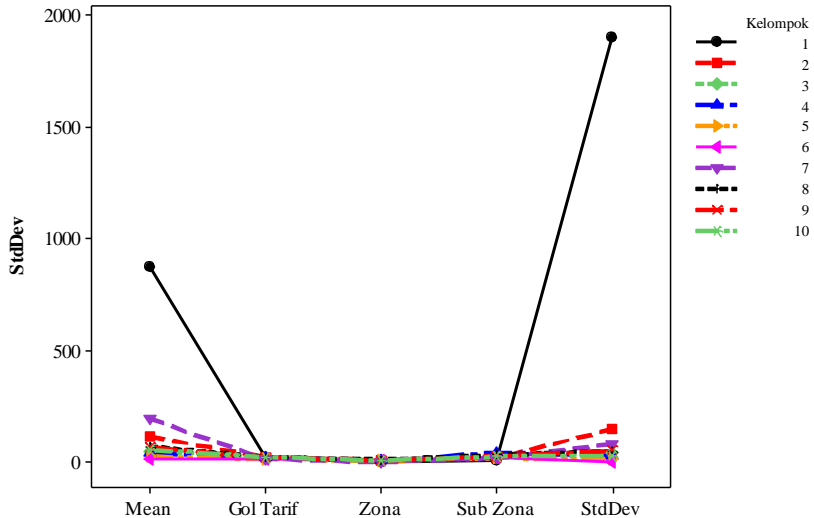
pada masing-masing kelompok yang terbentuk periode November 2017 hingga Februari 2018 pada Gambar 4.31 berikut.



**Gambar 4.31** Rata-rata Kelompok Periode November 2017 – Februari 2018

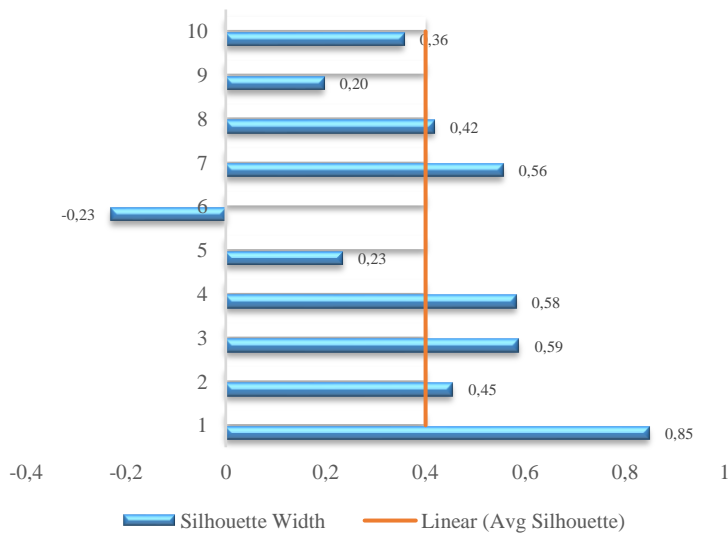
Secara graifik Gambar 4.31, kelompok yang terbentuk telah mempunyai rata-rata yang berbeda pada masing-masing variabel. Perbedaan rata-rata yang terlihat jelas ditunjukkan pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Gambar 4.31 juga menunjukkan kelompok 1 mempunyai rata-rata yang paling tinggi pada variabel rata-rata pemakaian air (*Mean*) dan disusul oleh kelompok 7. Selain variabel rata-rata pemakaian air, kelompok 1 juga memiliki nilai rata-rata yang tinggi pada variabel deviasi standar pemakaian air (*StdDev*) dan disusul oleh kelompok 9. Seperti halnya periode-periode sebelumnya, perbedaan rata-rata pada masing-masing kelompok hanya terlihat jelas pada variabel pemakaian rata-rata air dan deviasi standar pemakaian air. Selanjutnya disajikan Gambar 4.32 yang

menampilkan deviasi standar kelompok periode November 2017 hingga Februari 2018.



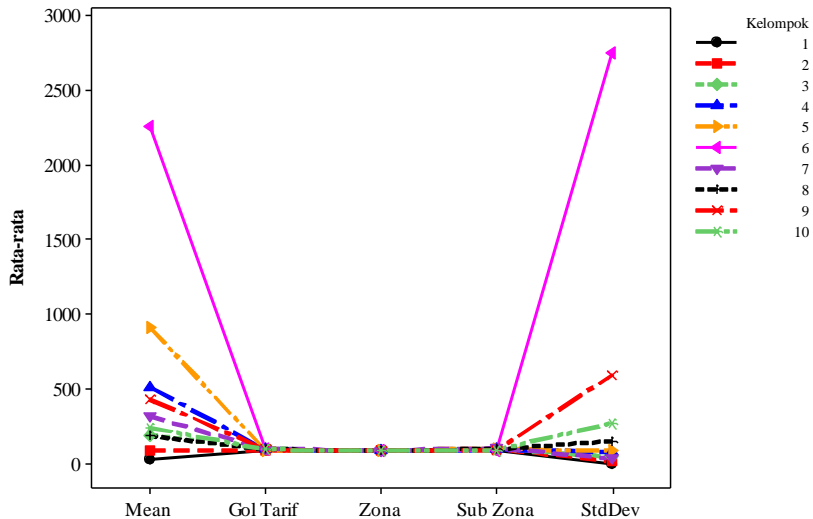
**Gambar 4.32** Deviasi Standar Kelompok Periode November 2017 – Februari 2018

Seperti halnya pada Gambar 4.30, Gambar 4.31 juga menunjukkan kelompok 1 mempunyai deviasi standar yang tinggi pada variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air. Gambar 4.32 juga menunjukkan perbedaan deviasi standar terlihat jelas pada variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air. Kelompok 10 juga mempunyai nilai deviasi standar tertinggi pada variabel rata-rata pemakaian air berdasarkan golongan tarif (*Gol Tarif*). Selain kelompok 1, kelompok 2 mempunyai deviasi standar yang tinggi pada variabel deviasi standar pemakaian air setelah kelompok 1.. Secara grafik kelompok yang terbentuk pada periode Agustus 2017 hingga Februari 2018 telah mempunyai deviasi standar yang berbeda. Berikutnya akan ditampilkan nilai *silhouette* dari periode Desember 2017 – Februari 2018 yang disajikan pada Gambar 4.33.



**Gambar 4.33** Nilai *Silhouette Width* 10 Klaster Periode Des 17–Feb 18

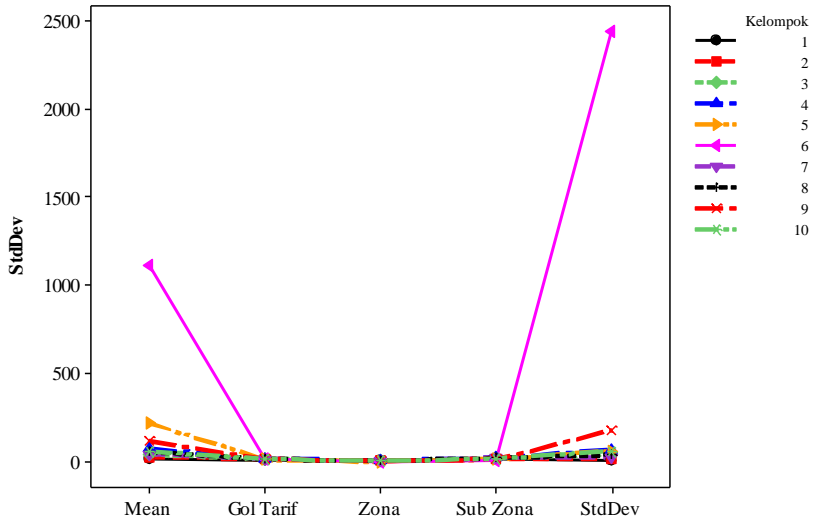
Gambar 4.33 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.40 yang berarti jika berdasarkan Tabel 2.1 kelompok yang lemah. Garis *orange* pada Gambar 4.33 merupakan nilai rata-rata yang digunakan sebagai batas penentuan suatu kelompok diduga sebagai kelompok anomali. Gambar 4.33 juga menunjukkan terdapat 4 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata. 3 dari 4 kelompok dibawah rata-rata tersebut masuk kedalam kriteria kelompok yang lemah karena nilai *silhouette width* kurang dari 0.5, sedangkan 1 kelompok yakni kelompok 6 mempunyai nilai *silhouette width* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Namun terdapat beberapa kelompok dengan nilai *silhouette width* lebih dari nilai rata-rata yang berada dalam kategori kelompok yang lemah. Selanjutnya akan disajikan Gambar 4.34 yang menjelaskan tentang rata-rata masing-masing kelompok yang terbentuk pada periode Desember 2017 hingga Februari 2018.



**Gambar 4.34** Rata-rata Kelompok Periode Desember 2017 – Februari 2018

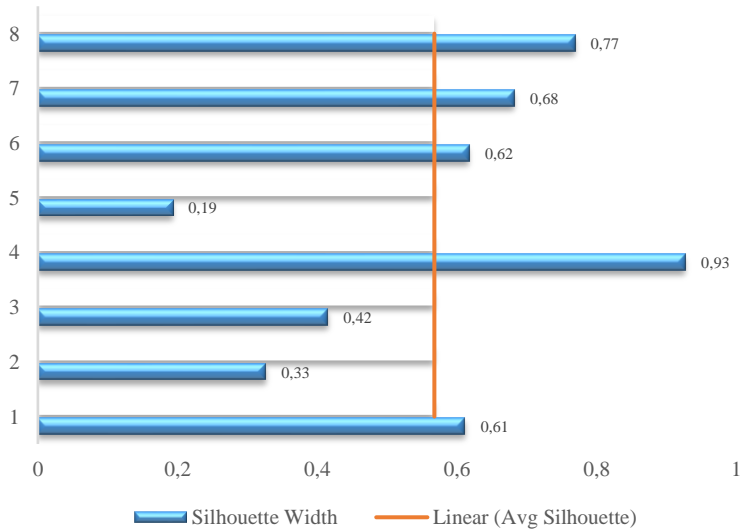
Gambar 4.34 menunjukkan nilai rata-rata masing-masing variabel pada periode Desember 2017 hingga Februari 2018. Rata-rata pada variabel rata-rata pemakaian air dan deviasi standar pemakaian air terdapat perbedaan pada masing-masing kelompok yang terbentuk. Pada variabel rata-rata pemakaian air, kelompok 6 mempunyai rata-rata pemakaian air paling tinggi dibandingkan kelompok lainnya dan kelompok 1 mempunyai nilai rata-rata yang paling rendah pada variabel rata-rata pemakaian air hingga deviasi standar. Perbedaan rata-rata pada variabel rata-rata pemakaian berdasarkan tarif (*Gol Tarif*), variabel rata-rata pemakaian air berdasarkan zona (*Zona*), dan variabel pemakaian air berdasarkan sub zona tidak begitu terlihat jelas seperti pada variabel rata-rata pemakaian air (*Mean*) dan deviasi standar pemakaian air (*StdDev*). Secara kelompok yang terbentuk sudah memiliki perbedaan karakteristik berdasarkan variabel-variabelnya. Selanjutnya akan

disajikan deviasi standar pada kelompok yang terbentuk periode Desember 2017 hingga Februari 2018 pada Gambar 4.35.



**Gambar 4. 35** Deviasi Standar Kelompok Periode Desember 2017 – Februari 2018

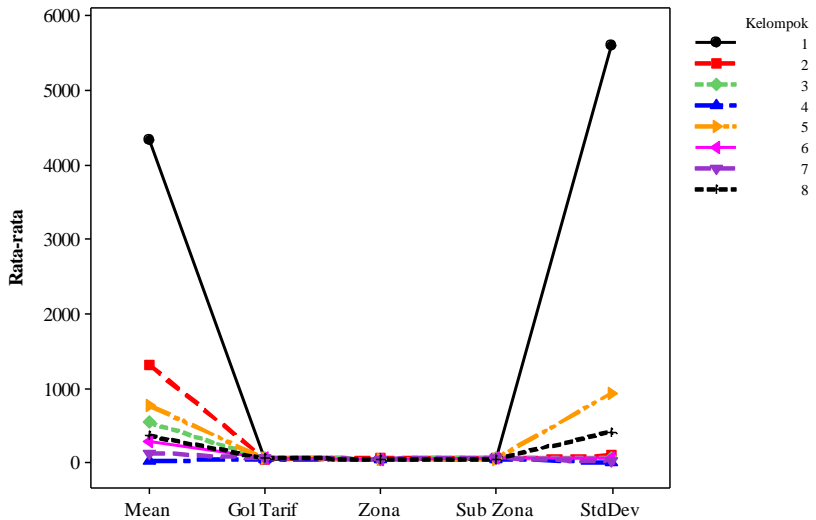
Gambar 4.35 menunjukkan terdapat perbedaan secara grafik deviasi standar pada masing-masing kelompok yang terbentuk pada periode Desember 2017 hingga Februari 2018. Jika dilihat pada deviasi standar, kelompok 6 mempunyai nilai deviasi standar yang paling tinggi dibandingkan kelompok lainnya pada variabel rata-rata pemakaian air (*Mean*) disusul dengan kelompok 5 dan kelompok 2. Kelompok 6 mempunyai deviasi standar yang paling tinggi pada 2 variabel, yaitu variabel rata-rata pemakaian air dan variabel deviasi standar. Gambar 4.35 juga menunjukkan bahwa kelompok yang terbentuk pada periode Desember 2017 hingga Februari 2018 mempunyai perbedaan karakteristik deviasi standar walaupun hanya dilihat berdasarkan grafik saja. Berikutnya akan ditampilkan nilai *silhouette* dari periode Januari – Februari 2018 yang disajikan pada Gambar 4.36.



**Gambar 4.36** Nilai *Silhouette Width* 8 Klaster Periode Jan – Feb 18

Gambar 4.36 menunjukkan rata-rata nilai *silhouette* adalah sebesar 0.56 yang berarti jika berdasarkan Tabel 2.1 kelompok yang cukup baik. Batas yang digunakan sebagai penentuan kelompok anomali adalah nilai rata-rata tersebut yang disimbolkan dengan garis berwarna *orange* pada Gambar 4.36. Selanjutnya Gambar 4.36 juga menunjukkan terdapat 3 kelompok yang mempunyai nilai *silhouette width* kurang dari *silhouette* rata-rata. 2 dari 3 kelompok dibawah rata-rata tersebut masuk kedalam kriteria kelompok yang lemah karena nilai *silhouette width* kurang dari 0.5, sedangkan 1 kelompok yakni kelompok 5 mempunyai nilai *silhouette width* kurang dari 0.25 yang berarti tidak dapat disebut sebagai kelompok. Selanjutnya disajikan Gambar 4.37 dan 4.38 yang membahas tentang karakteristik rata-rata dan deviasi standar tiap kelompok yang terbentuk pada periode Januari hingga Februari 2018.

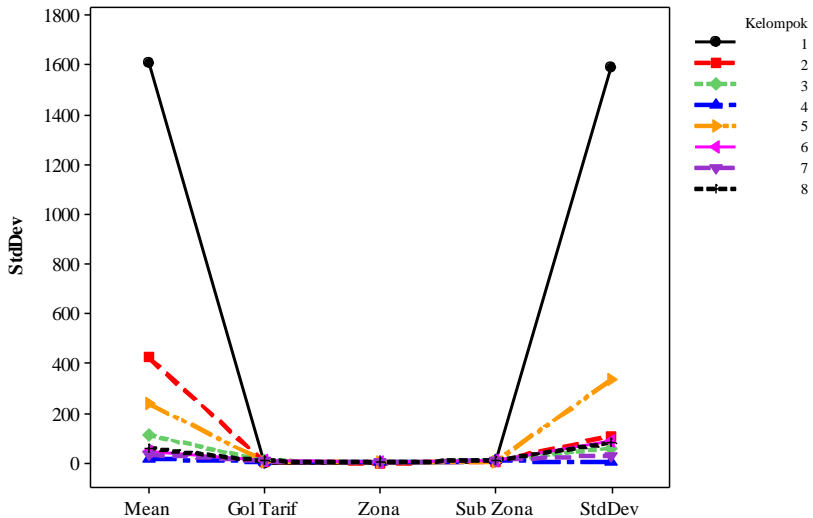




**Gambar 4.37** Rata-rata Kelompok Periode Januari – Februari 2018

Gambar 4.37 menunjukkan rata-rata kelompok pada periode Januari hingga Februari 2018 mengalami perbedaan yang jelas secara grafik pada dua variabel, yaitu variabel rata-rata pemakaian air (*Mean*) dan variabel deviasi standar pemakaian air (*StdDev*). Kelompok 1 mempunyai nilai rata-rata yang paling tinggi pada variabel rata-rata pemakaian air dan variabel deviasi standar pemakaian air. Setelah kelompok 1, kelompok 2 mempunyai nilai rata-rata yang tinggi pada variabel rata-rata pemakaian air. Variabel rata-rata pemakaian air berdasarkan golongan tarif, rata-rata pemakaian air berdasarkan zona, dan rata-rata pemakaian air berdasarkan sub zona tidak terdapat perbedaan yang sangat jelas secara grafik seperti dua variabel lainnya. Kelompok 4 mempunyai nilai rata-rata yang paling rendah dibandingkan kelompok lainnya pada variabel rata-rata pemakaian air. Secara grafik, kelompok yang terbentuk pada periode tersebut sudah berbeda dalam hal rata-rata, terutama pada variabel rata-rata pemakaian air dan variabel deviasi standar pemakaian air. Selanjutnya disajikan Gambar 4.38

yang membahas tentang deviasi standar pada tiap kelompok yang terbentuk pada periode Januari hingga Februari 2018.



**Gambar 4.38** Deviasi Standar Kelompok Periode Januari – Februari 2018

Gambar 4.38 menunjukkan adanya perbedaan deviasi standar sangat jelas secara grafik pada variabel rata-rata pemakaian air dan deviasi standar pemakaian air. Kelompok 1 mempunyai deviasi standar yang tinggi pada variabel rata-rata pemakaian air dan variabel deviasi standar pemakaian air. Hal tersebut disebabkan anggota didalam kelompok 1 mempunyai rata-rata pemakaian yang tinggi dibandingkan dengan kelompok lainnya dan mempengaruhi perhitungan deviasi standar pada variabel rata-rata pemakaian dan variabel deviasi standar pemakaian air. Setelah kelompok 1, kelompok 2 mempunyai nilai deviasi standar yang tinggi pada variabel rata-rata pemakaian air disusul oleh kelompok 5. Sebaliknya, kelompok 5 mempunyai deviasi standar yang tinggi pada variabel deviasi standar pemakaian air setelah kelompok 1 dan disusul oleh kelompok 2. Secara grafik kelompok yang

terbentuk telah memiliki perbedaan dalam hal deviasi standar. Setelah menghitung masing-masing nilai *silhouette* pada masing-masing periode selanjutnya adalah mendapatkan kelompok anomali.

Nilai *silhouette* menggambarkan sebaran data pengamatan, seperti pada Tabel 2.1, semakin besar nilai *silhouette* maka semakin bagus kelompok yang terbentuk. Berdasarkan Tabel 2.1 juga klaster yang dianggap sebagai kelompok anomali adalah klaster yang mempunyai nilai *silhouette width* kurang dari *silhouette width* rata-rata klaster tersebut, klaster-klaster tersebut disajikan pada Tabel 4.3 berikut.

**Tabel 4.3** Kelompok Anomali Masing-masing Periode

Periode	Klaster	Silhouette Width	Ukuran
Mar 17 - Feb 18	4	0.09	3
	2	0.31	322877
	5	0.32	223
	3	0.34	18
	6	0.35	20
Apr 17 - Feb 18	8	0.18	467
	3	0.32	130
	4	0.39	68
	7	0.39	21
	6	0.41	3
Mei 17 - Feb 18	3	-0.23	359
	6	0.48	23
	7	0.52	63
	2	0.55	25
Jun 17 - Feb 18	6	0.09	338
	3	0.36	366
	7	0.36	21
	2	0.42	12206
Jul 17 - Feb 18	5	0.21	2695
	9	-0.09	1722
	2	0.35	22
	4	0.51	565

**Tabel 4.3** Kelompok Anomali Masing-masing Periode (Lanjutan)

Periode	Klaster	Silhouette Width	Ukuran
Agt 17 - Feb 18	8	-0.19	2
	5	0.31	2
	2	0.56	152
	7	0.58	97
	9	1.00	1
Sep 17 - Feb 18	6	0.02	16
	4	0.26	12
	7	0.5	2
Okt 17 - Feb 18	2	0.14	6905
	6	0.33	75
	5	0.35	15
	3	0.42	1037
Nov 17 - Feb 18	1	-0.34	5
	2	0.25	23
	10	0.26	656
	3	0.32	398
Des 17 - Feb 18	5	0.23	19
	6	-0.23	4
	9	0.20	22
	10	0.36	76
Jan 18 - Feb 18	5	0.19	10
	2	0.33	8
	3	0.42	102

Tabel 4.3 menunjukkan 45 kelompok yang terduga anomali dari jumlah seluruh periode. Ukuran menjelaskan banyaknya anggota atau sampel dalam kelompok tersebut. Sebagian besar masing-masing periode terdapat 4 hingga 5 kelompok terduga anomali. Nilai *silhouette width* secara keseluruhan berada dibawah angka 0.25 dan dibawah 0.5. Artinya kelompok yang berada pada Tabel 4.3 merupakan kelompok yang lemah atau tidak dapat dikatakan sebagai kelompok karena persebaran anggotanya yang cukup menyebar atau dengan kata lain variansi yang besar dalam kelompok tersebut. Langkah berikutnya adalah menghitung nilai *outlier* untuk mengetahui anomali pada masing-masing kelompok tersebut. Perhitungan nilai *outlier* hanya dilakukan pada kelompok

yang mempunyai ukuran yang cukup besar. Kelompok yang dihitung *outlier score* adalah kelompok dengan ukuran lebih dari 100.

#### 4.5 Deteksi Anomali Menggunakan *Local Outlier Factor*

Setelah mendapatkan kelompok anomali menggunakan nilai *silhouette width* yang disajikan pada Tabel 4.4, selanjutnya adalah melakukan deteksi anomali pada klaster terpilih dengan ukuran klaster yang lebih dari 100. Hal tersebut dilakukan karena masih terlalu banyak pengamatan dalam menentukan apakah objek tersebut suatu anomali atau bukan.

Pada algoritma LOF ini, hal-hal yang diperhatikan adalah menentukan banyaknya nilai  $k$ , dengan  $k$  adalah jumlah observasi terdekat yang akan digunakan untuk menghitung nilai outlier. Kemudian menentukan nilai batas dalam memilih suatu anomali, penentuan batas tersebut dapat menggunakan kuantil dari data tersebut atau menentukan secara subjektif.

**Tabel 4.4** Hasil Deteksi Anomali Dengan Algoritma LOF

Periode	Klaster	Ukuran
Mar 17 - Feb 18	5	12
	2	16145
Apr 17 - Feb 18	8	23
	3	7
Mei 17 - Feb 18	3	18
	6	17
Jun 17 - Feb 18	3	19
	2	611
Jul 17 - Feb 18	5	135
	9	87
Agt 17 - Feb 18	4	29
	2	8
Okt 17 - Feb 18	2	346
	3	52
Nov 17 - Feb 18	10	33
	3	20

Tabel 4.4 menunjukkan jumlah anomali pada semua periode yang dianalisa menggunakan algoritma LOF, sebanyak 17562 kejadian terduga anomali. Nilai *threshold* didapatkan dari kuantil data hasil *Local Outlier Factor*. Dengan menggunakan batas *threshold* berdasarkan kuantil 95% didapatkan sebanyak 17562 kejadian anomali. Pada periode Maret 2017 hingga Februari 2018 merupakan periode penuh penelitian dengan dugaan anomali terbanyak pula. Karena ukuran klaster yang begitu besar sehingga dugaan anomalnya juga cukup besar. Nilai *threshold* yang bermacam-macam disebabkan nilai *outlier* serta jumlah data.

Data anomali yang didapatkan menggunakan algoritma SOM selanjutnya akan digabung dengan data hasil anomali menggunakan algoritma LOF, dan hasil tersebut merupakan hasil deteksi anomali pelanggan PDAM Surya Sembada Kota Surabaya. Berikut adalah hasil deteksi anomali yang disajikan pada Tabel 4.5 dibawah ini.

**Tabel 4.5** Jumlah Pelanggan Terdeteksi Anomali Per-Periode

Periode	Ukuran
Mar 17 - Feb 18	16198
Apr 17 - Feb 18	122
Mei 17 - Feb 18	129
Jun 17 - Feb 18	668
Jul 17 - Feb 18	273
Agt 17 - Feb 18	110
Sep 17 - Feb 18	30
Okt 17 - Feb 18	488
Nov 17 - Feb 18	124
Des 17 - Feb 18	121
Jan 18 - Feb 18	120

Tabel 4.5 menunjukkan jumlah keseluruhan terdeteksi anomali ada sebanyak 18383 kejadian. Tabel 4.5 menunjukkan

anomali terbanyak berada pada periode Maret 2017 hingga Februari 2018, sedangkan anomali yang terkecil adalah pada periode September 2017 hingga Februari 2018. Anomali yang didapatkan pada Tabel 4.5 merupakan data dari pengguna PDAM Surya Sembada Kota Surabaya, tentu saja terdapat kode identitas atau nomor pelanggan. Untuk mengetahui pelanggan yang terduga anomali adalah menghitung jumlah frekuensi pelanggan tersebut terdeteksi anomali pada 11 periode seperti pada Tabel 4.5, berikut akan disajikan Tabel 4.6 yang merupakan jumlah pelanggan terdeteksi anomali berdasarkan frekuensi.

**Tabel 4.6** Jumlah Pelanggan Terdeteksi Anomali Berdasarkan Frekuensi

<u>Frekuensi Terdeteksi</u>	<u>Jumlah</u>
1	16836
2	248
3	161
4	71
5	12
6	8
7	6
8	2
9	12
10	1

Tabel 4.6 menunjukkan sebagian besar pelanggan terdeteksi anomali hanya 1 periode saja. Berarti 16836 pelanggan tersebut masih belum mampu dikatakan sebagai anomali atau yang mempunyai konsumsi air yang tidak normal. Sedangkan pelanggan yang terdeteksi lebih dari 1 kali ada sebanyak 521 pelanggan. Sehingga hasil deteksi anomali dari 18383 kejadian didapatkan 17357 579 pelanggan yang terduga anomali. Namun tidak semua pelanggan terdeteksi dengan frekuensi lebih dari satu. Pelanggan yang dipilih sebagai pelanggan anomali adalah yang terdeteksi dengan frekuensi lebih dari 4. Hasilnya dapat dilihat pada Tabel 4.7 berikut.

**Tabel 4.7** Pelanggan Dengan Frekuensi Tertinggi

No Pelanggan	Frekuensi	No Pelanggan	Frekuensi
1017865	5	2150919	5
1153966	6	2153089	5
1166604	6	2202996	5
2016089	9	3020691	7
2016090	8	3071164	5
2016092	8	3120338	7
2016093	9	3120430	9
2016095	9	3120944	7
2080655	7	3133697	5
2136371	9	3133698	6
No Pelanggan	Frekuensi	No Pelanggan	Frekuensi
3134197	7	4361923	9
3134199	6	5040069	9
4010624	5	5063225	9
4010711	5	5105905	9
4082622	5	5110599	6
4098338	10	5110637	6
4100874	5	5111279	6
4105776	9	5120886	6
4123211	5	5404657	9
4280618	9	5426971	7
		5451698	5

Penelitian ini mempunyai 11 periode, atau 12 bulan waktu pengamatan. Tabel 4.7 menggunakan batas 5 periode atau pelanggan yang terdeteksi sebanyak 5 periode dianggap sebagai anomali. Pelanggan-pelanggan tersebut dicurigai mempunyai pola konsumsi yang tidak normal dibandingkan pelanggan lainnya. Jika dianggap batas 5 frekuensi terlalu jauh dari jumlah periodenya, hal tersebut dikarenakan anomali konsumsi air sifatnya singkat atau tidak begitu lama. PDAM Surya Sembada Kota Surabaya sendiri mempunyai cara sederhana untuk mengetahui apakah pelanggan tersebut merupakan anomalia tau sebaliknya. Yaitu dengan menghitung selisih pemakaian air bulan sekarang dengan bulan sebelumnya. Jika selisih tersebut lebih dari 50% pelanggan tersebut



terindikasi anomali. Namun juga harus didukung dengan data-data lainnya. Sehingga berdasarkan hal tersebut pada penelitian ini batas frekuensi yang digunakan untuk menentukan pelanggan tersebut anomali adalah sebanyak 5 kali.

#### 4.6 Hasil Deteksi Anomali Dengan Metode PDAM

Perhitungan dengan metode PDAM dilakukan untuk mengetahui seberapa baik model atau untuk membandingkan perbedaan hasil deteksi anomali antara metode *Kohonen-SOM* dan *Local Outlier Factor* dengan metode PDAM. Perhitungan dilakukan jika selama periode penelitian terhadap selisih yang lebih dari 50% maka dikatakan pelanggan tersebut terindikasi anomali. Data yang digunakan hanya 42 pelanggan yang terpilih berdasarkan metode *Kohonen-SOM* dan *Local Outlier Factor*. Berikut adalah hasil deteksi anomali sederhana yang digunakan oleh PDAM Surya Sembada Kota Surabaya.

**Tabel 4.8** Hasil Deteksi Anomali Menggunakan Metode PDAM Surya Sembada Kota Surabaya

No Pelanggan	Kategori	No Pelanggan	Kategori
1017865	Bukan Anomali	2150919	Bukan Anomali
1153966	Anomali	2153089	Bukan Anomali
1166604	Anomali	2202996	Anomali
2016089	Bukan Anomali	3020691	Bukan Anomali
2016090	Bukan Anomali	3071164	Bukan Anomali
2016092	Bukan Anomali	3120338	Anomali
2016093	Bukan Anomali	3120430	Anomali
2016095	Bukan Anomali	3120944	Bukan Anomali
2080655	Bukan Anomali	3133697	Bukan Anomali
2136371	Bukan Anomali	3133698	Bukan Anomali
3134197	Bukan Anomali	4361923	Anomali
3134199	Anomali	5040069	Bukan Anomali
4010624	Anomali	5063225	Anomali
4010711	Anomali	5105905	Bukan Anomali
4082622	Anomali	5110599	Anomali
4098338	Anomali	5110637	Bukan Anomali

**Tabel 4.8** Hasil Deteksi Anomali Menggunakan Metode PDAM Surya Sembada Kota Surabaya (Lanjutan)

No Pelanggan	Kategori	No Pelanggan	Kategori
4100874	Bukan Anomali	5111279	Anomali
4105776	Bukan Anomali	5120886	Bukan Anomali
4123211	Bukan Anomali	5404657	Bukan Anomali
4280618	Bukan Anomali	5426971	Anomali
		5451698	Anomali

Tabel 4.8 menunjukkan terdapat 42 pelanggan yang dideteksi anomali menggunakan metode *Kohonen-SOM* dan *Local Outlier Factor*. Setelah dilakukan perhitungan selisih, didapatkan hanya 16 pelanggan yang terdeteksi anomali dan 25 lainnya bukan anomali. Jika dilakukan perhitungan akurasi prediksi, ketepatan model dalam mendeteksi anomali pelanggan PDAM Surya Sembada Kota Surabaya sebesar 39%. Sehingga perlu diketahui bagaimana karakteristik pelanggan yang dideteksi anomali dengan metode *Kohonen-SOM* dan *Local Outlier Factor* namun tidak terdeteksi oleh metode sederhana PDAM Surya Sembada Kota Surabaya.

#### 4.7 Karakteristik Pelanggan Yang Terdeteksi Anomali

Pada analisis yang telah dilakukan terhadap 454160 pelanggan PDAM Surya Sembada Kota Surabaya, terdapat 41 pelanggan yang terdeteksi mempunyai pola konsumsi yang tidak normal. Namun dengan menggunakan metode PDAM Surya Sembada Kota Surabaya hanya terdeteksi 16 dari 41 pelanggan tersebut. Tabel 4.9 disajikan untuk melihat karakteristik pelanggan yang terdeteksi anomali.

**Tabel 4.9** Karakteristik Pelanggan Terdeteksi Anomali

No Pelanggan	Rata-rata	Rata-rata Golongan	Rata-rata Sub-Zona	Standar Deviasi
1017865	683.67	335.69	351.68	67.11
1153966	927.75	361.39	352.33	2939.80
1166604	542.83	361.39	323.46	1834.80
2016089	1087.00	335.69	380.83	85.82
2016090	844.50	335.69	380.83	162.99
2016092	906.67	335.69	380.83	59.40

**Tabel 4.9** Karakteristik Pelanggan Terdeteksi Anomali (lanjutan)

<b>No Pelanggan</b>	<b>Rata-rata</b>	<b>Rata-rata Golongan</b>	<b>Rata-rata Sub-Zona</b>	<b>Standar Deviasi</b>
2016093	1286.67	335.69	380.83	176.49
2016095	947.67	335.69	380.83	70.34
2080655	750.08	487.02	484.54	186.22
2136371	1086.42	335.69	355.28	269.77
2150919	187.50	361.39	463.85	461.62
2153089	675.17	487.02	463.85	75.55
2202996	255.92	324.81	315.00	748.81
3020691	682.08	487.02	386.78	169.78
3071164	148.08	342.53	301.91	454.72
3120338	436.83	342.53	278.58	148.51
3120430	783.00	342.53	278.58	334.50
3120944	500.00	342.53	278.58	0.00
3133697	466.83	324.81	299.90	76.13
3133698	661.58	324.81	299.90	171.86
3134197	696.92	324.81	299.90	166.59
3134199	578.75	324.81	299.90	251.40
4010624	629.75	487.02	508.10	561.11
4010711	619.67	487.02	508.10	225.72
4082622	443.67	361.39	401.65	672.06
4098338	1097.92	335.69	371.40	490.07
4100874	663.25	324.81	420.91	74.58
4105776	952.17	361.39	332.23	181.74
4123211	504.17	361.39	412.77	171.55
4280618	1216.08	335.69	483.65	72.95
4361923	755.25	335.69	286.67	153.22
5040069	2025.00	487.02	387.87	296.44
5063225	680.42	487.02	448.04	289.51
5105905	994.58	324.81	426.11	141.63
5110599	138.67	487.02	671.96	220.83
5110637	469.67	487.02	671.96	30.85
5111279	145.33	487.02	671.96	192.23
5120886	154.25	487.02	557.53	171.91
5404657	1093.17	361.39	433.95	116.19
5426971	222.92	324.81	278.39	171.15
5451698	187.50	487.02	562.63	291.54

Tabel 4.9 menunjukkan bahwa pelanggan yang dideteksi sebagai anomali dengan metode *Kohonen-SOM* dan *Local Outlier Factor* adalah pelanggan yang memiliki rata-rata pemakaian air melebihi 50% rata-rata pemakaian air pada golongan dan sub-zona pelanggan tersebut. Seperti pada nomor pelanggan 1017865 tidak dideteksi anomali pada Tabel 4.8, namun pada dapat dilihat karakteristiknya pada Tabel 4.9, nilai rata-rata konsumsi airnya melebihi rata-rata pemakaian golongan dan sub-zonanya. Kemudian pada nomor pelanggan 3120944, dengan metode deteksi anomali dari PDAM tidak dikategorikan sebagai anomali karena tidak terdapat selisih yang lebih dari 50% dari bulan sekarang dengan bulan sebelumnya. Namun pada metode *Kohonen-SOM* dan *Local Outlier Factor* pelanggan tersebut dideteksi sebagai anomali, karena mempunyai pola konsumsi yang sama setiap bulannya sehingga menyebabkan deviasi standar pemakaian airnya adalah 0. Mungkin hal tersebut perlu dilakukan konfirmasi pada pihak PDAM Surya Sembada Kota Surabaya. Setelah mengetahui karakteristik pelanggan yang terdeteksi anomali, selanjutnya adalah menganalisa pelanggan yang tidak terdeteksi anomali menggunakan metode sederhana PDAM Surya Sembada Kota Surabaya dengan *Kohonen-SOM* dan *Local Outlier Factor*, hasilnya dapat dilihat pada Tabel 4.10.

**Tabel 4.10** Pelanggan Yang Tidak Terdeteksi Anomali Dengan Metode Sederhana PDAM

No Urut	No Plg	Mean	Maks	Mean Gol	Mean Sub-Zona	Std Dev
1	1017865	683.67	788	335.69	351.68	67.11
2	2016089	1087.00	1235	335.69	380.83	85.82
3	2016090	844.50	989	335.69	380.83	162.99
4	2016092	906.67	1005	335.69	380.83	59.40
5	2016093	1286.67	1606	335.69	380.83	176.49
6	2016095	947.67	1072	335.69	380.83	70.34
7	2080655	750.08	984	487.02	484.54	186.22

**Tabel 4.10** Pelanggan Yang Tidak Terdeteksi Anomali Dengan Metode Sederhana PDAM (lanjutan)

<b>No Urut</b>	<b>No Plg</b>	<b>Mean</b>	<b>Maks</b>	<b>Mean Gol</b>	<b>Mean Sub-Zona</b>	<b>Std Dev</b>
8	2136371	1086.42	1516	335.69	355.28	269.77
9	2153089	675.17	757	487.02	463.85	75.55
10	3020691	682.08	1001	487.02	386.78	169.78
11	3133697	466.83	661	324.81	299.90	76.13
12	3133698	661.58	976	324.81	299.90	171.86
13	3134197	696.92	1095	324.81	299.90	166.59
14	4100874	663.25	785	324.81	420.91	74.58
15	4105776	952.17	1110	361.39	332.23	181.74
16	4123211	504.17	739	361.39	412.77	171.55
17	4280618	1216.08	1304	335.69	483.65	72.95
18	5040069	2025.00	2515	487.02	387.87	296.44
19	5105905	994.58	1203	324.81	426.11	141.63
20	5404657	1093.17	1235	361.39	433.95	116.19
21	3071164	148.08	1592	342.53	301.91	454.72
22	5110637	469.67	548	487.02	671.96	30.85
23	5120886	154.25	554	487.02	557.53	171.91
24	3071164	148.08	1592	342.53	301.91	454.72
25	3120944	500.00	500	342.53	278.58	0.00

Pelanggan dengan nomor urut 1 hingga 21 adalah pelanggan yang mempunyai pola rata-rata pemakaian air yang lebih dari rata-rata pemakaian air golongan dan sub-zonanya. Sehingga oleh Kohonen-SOM dan LOF dianggap sebagai anomali. Sedangkan berdasarkan metode sederhana PDAM tidak mampu menangkap hal tersebut karena hanya menggunakan selisih antar bulan. Pelanggan dengan no urut 22 hingga 24 mempunyai karakteristik sebaliknya, yaitu nilai rata-rata pemakaian airnya lebih kecil dibandingkan rata-rata pemakaian air golongan dan sub-zonanya.

Serta mempunyai karakteristik nilai pemakaian maksimum yang sangat tinggi disbanding-kan rata-rata pemakaian setiap bulannya. Sedangkan pelanggan dengan no urut 25 adalah mempunyai pola konsumsi yang sama setiap bulannya serta berada lebih dari rata-rata pemakaian air golongan dan subzonanya. Pemakaian yang sama menyebabkan standar deviasi pelanggan tersebut nol. Sehingga oleh Kohonen SOM dan LOF dianggap sebagai perilaku konsumsi yang tidak normal.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Penelitian deteksi anomali pada pola konsumsi pelanggan PDAM di Kota Surabaya menghasilkan kesimpulan sebagai berikut :

1. Hasil deteksi anomali menggunakan metode *Kohonen SOM* dan *local outlier factor* adalah sebanyak 41 pelanggan. Sedangkan dengan menggunakan metode sederhana PDAM hanya didapatkan 16 pelanggan dari 41 pelanggan yang terdeteksi menggunakan metode *Kohonen SOM* dan *local Outlier Factor*. Metode *Kohonen SOM* dan *Local Outlier Factor* mampu menangkap pola anomali yang tidak dapat ditangkap dengan menggunakan metode dari PDAM. Dari 41 pelanggan tersebut 1 pelanggan terdeteksi lebih dari 8 kali dalam periode penelitian.
2. Karakteristik pelanggan yang terdeteksi anomali rata-rata pemakaian airnya berbeda jauh dari rata-rata pemakaian air di zona dan sub zona sekitarnya. Selain itu karakteristik pola pemakaian konsumsi air pelanggan anomali berbeda jauh dari pelanggan yang normal.

#### **5.2 Saran**

Berdasarkan penelitian yang telah dilakukan, saran yang dapat diberikan untuk PDAM Surya Sembada Kota Surabaya adalah melakukan investigasi dengan kurun waktu yang optimal. Karena pada data selama periode Maret 2017 hingga Februari 2018 terdapat banyak sekali pola-pola pelanggan yang diduga anomali. Seperti nilai catat meter yang selalu sama tiap bulannya, kemudian tinggi dan rendah pada bulan tertentu dengan selisih lebih dari 50% dari pemakaian sebelumnya. Selain itu banyak sekali pelanggan dengan angka meteran yang sama persis setiap bulannya dalam periode tersebut. Sedangkan saran untuk pengembangan penelitian selanjutnya adalah, pilih salah satu wilayah yang ingin dijadikan objek penelitian. Salah satu wilayah tersebut dapat ditentukan berdasarkan pembagian Zona PDAM Surya Sembada Kota

Surabaya. Karena jika terlalu banyak informasi atau terlalu banyak pengamatan membutuhkan proses yang lama untuk membangun model. Kemudian lakukan *Pre-processing* sebaik mungkin untuk hasil yang lebih maksimal. Selain itu, jika menggunakan metode *Kohonen-SOM*, hindari *grid* yang kosong atau tidak terdapat observasi didalamnya. Karena hal tersebut dapat meningkatkan tingkat kebaikan klaster. Persiapkan perangkat olah data yang memadai karena data yang digunakan cukup besar.



## DAFTAR PUSTAKA

- Agaskar, V., Babariya, M., Chandran, S., & Giri, N. (2017). Unsupervised Learning for Credit Card Fraud Detection. *International Research Journal of Engineering and Technology (IRJET)*, 2343-2346.
- Aggarwal, C. C. (2013). *Outlier Analysis*. Boston: Kluwer Academy Publisher.
- Ahmad, P. H., & Dang, S. (2015). Performance Evaluation Clustering Algorithm Using Different Datasets. *International Journal of Advance Research in Computer Science and Management Studies*, 167-173.
- Amorim, & Henning. (2015). Recovering The Number of Clusters in Data Sets With Noise Using Feature Rescaling Factors. *Information Sciences*, 126-145.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. *To Appear in ACM Computing Surveys*, 1-72.
- Choudhary, P. (2017, 02 14). *DataScience.com*. Diambil kembali dari Data Science: [www.datascience.com/blog/python-anomaly-detection](http://www.datascience.com/blog/python-anomaly-detection)
- Dang, X. H., Micenkova, B., Assent, I., & Ng, R. T. (2013). Local Outlier Detection With Interpretation. *ECML*. Prague: PPKD.
- Domingues, R. (2015). *Machine Learning for Unsupervised Fraud Detection*. Stockholm: KTH Royal Institute of Technology.
- dos, A. E., Saavedra, O., Cortes, O., & de Souza, A. (2011). Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution System. *IEEE Transaction on Power Delivery*, 2436-2442.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (t.thn.). A Density-Based Algorithm for Discovering Clusters in Large Spatial Database With Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226-231.

- Furqon, M. T., & Muflikhah, L. (2016). Clustering The Potential Risk Of Tsunami Using Density-Based Spatial Clustering of Application With Noise (DBSCAN). *Journal Environmental Engineering & Suistainable Technology*, 1-8.
- Goldstein, M., & Uchida, S. (2016). A Comperative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLoS ONE*, 1-31.
- Hair, Anderson, Tatham, & Black. (2007). *Multivariate Data Analysis*. New Jersey: Prentice Hall International Inc.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques (3rd ed.)*. Waltham: Morgan Kaufmann.
- Hodge, V., & Austin, J. (2004). A Survey of Outlier Detection Methodologist. *Artificial Intelligence Review* , 85-126.
- Jannah, S. Z., & Fithriasari, K. (2015). *Deteksi Anomali Pada Konsumsi Listrik Pelanggan Area Pelayanan Surabaya Selatan Menggunakan Algoritma Kohonen Self-Organizing Maps*. Surabaya: Departemen Statistika Institut Teknologi Sepuluh Nopember.
- Jawa Pos. (2015, 05 4). *Jawa Pos*. Diambil kembali dari jawapos.com:  
<http://www2.jawapos.com/baca/artikel/16741/layanan-pdam-surya-sembada-meluas-pencurian-air-ikut-naik>
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Johnson, R., & Winchern, D. (2007). *Applied Multivariate Statistical Analysis (6th ed.)*. New Jersey: Person Prentice Hall.
- JPNN. (2016, 8 19). *JPNN*. Diambil kembali dari JPNN.com:  
<https://www.jpnn.com/news/pdam-sikat-para-pencuri-air-masyarakat>
- Kaiser, H. (1974). An Index Of Factor Simplicity. *Psychometrika* 39, 31-36.
- Kohonen, T. (2001). *Self-Organizing Maps*. Heidelberg: Springer.

- Lapisdou, C., Papageorgiu, E., Kokkinos, K., Sahu, S., Gupta, A., & Tassiulas, L. (2015). Exploring Patterns in Water Consumption by Clustering. *13th Computer Control for Water Industry Conference, CCWI 2015* (hal. 1439-1446). Amsterdam: Elsevier:Procedia Engineering.
- Mulyadi, A., Bijaksana, M. A., & Kusumo, D. S. (2007). Analisis dan Implementasi Metode Local Outlier Factor Untuk Mendeteksi Anomali Data. 1-38.
- Nian, K., Zhang, H., Tayal, A., Coleman, T., & Li, Y. (2016). Auto Insurance Fraud Detection Using Unsupervised Spectral Ranking for Anomaly. *The Journal of Finance and Data Science*, 58-75.
- Omar, S., Ngadi, A., & Jebur, H. H. (2013). Machine Learning Techniques for Anomaly Detection: An Overview. *International Journal of Computer Application*, 33-41.
- Zakariya, A. Z., Djunaidy, A., & Kusumawardani, R. P. (2012). Pembuatan Aplikasi Pendeteksi Anomali pada Pola Konsumsi Listrik Pelanggan Kota Surabaya Menggunakan Algoritma Klasterisasi Berbasis Densitas. Tugas Akhir Sistem Informasi. *Jurnal Teknik POMITS Vol. 1*, 1-5.

*(Halaman sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1 Data Pemakaian Air PDAM Surya Sembada Kota Surabaya Periode Maret 2017-Februari 2018

No mor	No PLG	Zo na	Sub Zona	Gol Tarif	Mar -17	Apr -17	May -17	Jun -17	Jul -17	Aug -17	Sep -17	Oct -17	Nov -17	Dec -17	Jan -18	Feb -18
1	4	0	0	4B.2	8	9	10	10	7	17	60	81	113	120	79	93
2	8	0	0	4B.2	0	2	4	3	1	3	3	2	1	3	2	11
3	9	0	0	4B.2	37	38	38	43	59	45	39	41	42	33	27	31
4	10	0	0	4B.2	32	34	36	12	10	0	33	41	26	35	28	38
5	12	0	0	4B.2	4	9	9	11	11	10	21	14	10	16	7	5
6	13	0	0	4B.2	43	30	28	31	24	38	42	49	27	33	41	22
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
500 761	5460 577	5	546	3C.2	0	0	0	0	0	0	0	7	1	4	1	0
500 762	5460 579	5	546	3C.2	25	43	34	26	31	29	40	39	41	23	9	15
500 763	5460 580	5	546	3C.2	24	29	26	25	25	33	41	26	9	1	0	1
500 764	5470 001	5	542	3A	37	43	47	49	33	44	33	32	25	30	27	50

**Lampiran 2** Data Hasil *Pre-Processing* Pemakaian Air Pelanggan PDAM Surya Sembada Kota Surabaya

**1. Data 12 Bulan (Maret 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	50.58333	120	361.3946	331.1141	326.5205	44.92308
2	2.916667	11	361.3946	331.1141	326.5205	2.778434
3	39.41667	59	361.3946	331.1141	326.5205	8.050503
4	27.08333	41	361.3946	331.1141	326.5205	12.85201
5	10.58333	21	361.3946	331.1141	326.5205	4.699291
6	34	49	361.3946	331.1141	326.5205	8.474561
7	35.16667	47	361.3946	331.1141	326.5205	8.922223
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	29.58333	43	487.0175	363.0814	608.6934	10.62979
454155	20	41	487.0175	363.0814	608.6934	13.72456
454156	37.5	50	324.8054	363.0814	278.3893	8.764598
454157	19.33333	24	342.5307	362.6595	269.5088	2.46183
454158	28.75	35	324.8054	285.9797	252.4164	3.646293
454159	38.08333	52	324.8054	363.0814	332.2335	6.459079
454160	16.91667	23	324.8054	363.0814	311.9326	2.151462

**2. Data 11 Bulan (April 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	54.45455	120	332.769	305.2338	301.086	44.96746
2	3.181818	11	332.769	305.2338	301.086	2.750207
3	39.63636	59	332.769	305.2338	301.086	8.405626
4	26.63636	41	332.769	305.2338	301.086	13.38113
5	11.18182	21	332.769	305.2338	301.086	4.42308
6	33.18182	49	332.769	305.2338	301.086	8.376374
7	36	47	332.769	305.2338	301.086	8.854377
8	6.090909	12	332.769	305.2338	301.086	2.700168
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	30	43	448.354	334.462	564.146	11.04536
454155	19.63636	41	448.354	334.462	564.146	14.33369
454156	37.54545	50	299.2719	334.462	256.3168	9.190905
454157	19.54545	24	315.5889	333.6182	248.0093	2.46429
454158	28.90909	35	299.2719	263.4664	232.1995	3.780332
454159	38	52	299.2719	334.462	305.8282	6.76757
454160	16.81818	23	299.2719	334.462	287.1438	2.227922

### 3. **Data 10 Bulan (Mei 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	59	120	303.5333	277.8732	274.165	44.65672
2	3.3	11	303.5333	277.8732	274.165	2.869379
3	39.8	59	303.5333	277.8732	274.165	8.84182
4	25.9	41	303.5333	277.8732	274.165	13.86803
5	11.4	21	303.5333	277.8732	274.165	4.599517
6	33.5	49	303.5333	277.8732	274.165	8.759122
7	36.3	47	303.5333	277.8732	274.165	9.274218
8	5.5	9	303.5333	277.8732	274.165	1.95789
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	28.7	41	408.6546	304.7152	519.9708	10.71914
454155	18.7	41	408.6546	304.7152	519.9708	14.75014
454156	37	50	272.8879	304.7152	233.7298	9.498538
454157	19.6	24	287.8002	303.4205	224.813	2.590581
454158	28.7	35	272.8879	239.896	210.3349	3.917199
454159	37.5	52	272.8879	304.7152	279.0251	6.916165
454160	16.9	23	272.8879	304.7152	261.8714	2.330951



**4. Data 9 Bulan (Juni 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	64.44444	120	274.0299	250.0127	246.7645	43.70386
2	3.222222	11	274.0299	250.0127	246.7645	3.032234
3	40	59	274.0299	250.0127	246.7645	9.354143
4	24.77778	41	274.0299	250.0127	246.7645	14.21951
5	11.66667	21	274.0299	250.0127	246.7645	4.795832
6	34.11111	49	274.0299	250.0127	246.7645	9.061518
7	35.55556	47	274.0299	250.0127	246.7645	9.514609
8	5.555556	9	274.0299	250.0127	246.7645	2.068279
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	28.11111	41	368.9237	275.7549	473.7299	11.19648
454155	17.88889	41	368.9237	275.7549	473.7299	15.40653
454156	35.88889	50	246.4021	275.7549	210.8201	9.360081
454157	20	24	259.9215	273.7408	204.8543	2.397916
454158	29.11111	35	246.4021	216.3599	189.0346	3.919325
454159	37.22222	52	246.4021	275.7549	252.1066	7.276293
454160	16.88889	23	246.4021	275.7549	235.9231	2.472066

**5. Data 8 Bulan (Juli 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	71.25	120	243.9294	222.5662	219.6742	41.30981
2	3.25	11	243.9294	222.5662	219.6742	3.24037
3	39.625	59	243.9294	222.5662	219.6742	9.927415
4	26.375	41	243.9294	222.5662	219.6742	14.31221
5	11.75	21	243.9294	222.5662	219.6742	5.119989
6	34.5	49	243.9294	222.5662	219.6742	9.606545
7	35.375	47	243.9294	222.5662	219.6742	10.15505
8	5.75	9	243.9294	222.5662	219.6742	2.12132
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	28.375	41	328.3078	245.7709	423.4672	11.93958
454155	17	41	328.3078	245.7709	423.4672	16.22168
454156	34.25	50	219.3878	245.7709	186.9582	8.514693
454157	20.125	24	231.4242	243.4946	183.2169	2.531939
454158	29.5	35	219.3878	192.5522	167.779	4
454159	37.5	52	219.3878	245.7709	224.3665	7.727502
454160	16.75	23	219.3878	245.7709	209.3973	2.60494

**6. Data 7 Bulan (Agustus 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	80.42857	120	216.7928	196.4887	194.0395	34.70762
2	3.571429	11	216.7928	196.4887	194.0395	3.359422
3	36.85714	45	216.7928	196.4887	194.0395	6.593648
4	28.71429	41	216.7928	196.4887	194.0395	13.70784
5	11.85714	21	216.7928	196.4887	194.0395	5.520524
6	36	49	216.7928	196.4887	194.0395	9.309493
7	35.28571	47	216.7928	196.4887	194.0395	10.96531
8	5.571429	9	216.7928	196.4887	194.0395	2.225395
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	28	41	292.0449	219.772	383.9781	12.84523
454155	15.85714	41	292.0449	219.772	383.9781	17.17002
454156	34.42857	50	195.0773	219.772	167.5744	9.180725
454157	20.85714	24	206.12	217.072	166.0018	1.573592
454158	28.71429	32	195.0773	171.3706	146.5824	3.59232
454159	38.28571	52	195.0773	219.772	197.5878	7.994045
454160	16.85714	23	195.0773	219.772	188.3727	2.794553

**7. Data 6 Bulan (September 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	91	120	185.4104	168.2296	166.1185	22.51222
2	3.666667	11	185.4104	168.2296	166.1185	3.669696
3	35.5	42	185.4104	168.2296	166.1185	6.058052
4	33.5	41	185.4104	168.2296	166.1185	5.75326
5	12.16667	21	185.4104	168.2296	166.1185	5.980524
6	35.66667	49	185.4104	168.2296	166.1185	10.15218
7	33.33333	47	185.4104	168.2296	166.1185	10.5956
8	91	120	185.4104	168.2296	166.1185	22.51222
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	33.66667	37	250.9968	187.2911	332.5912	2.160247
454155	27.83333	41	250.9968	187.2911	332.5912	14.06295
454156	13	41	250.9968	187.2911	332.5912	16.88787
454157	32.83333	50	166.1927	187.2911	141.9188	8.931219
454158	20.33333	21	175.5349	184.9066	139.228	0.816497
454159	28.5	32	166.1927	145.87	126.8945	3.885872
454160	39.83333	52	166.1927	187.2911	171.2787	7.521081

**8. Data 5 Bulan (Oktober 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	97.2	120	154.0777	141.2592	139.519	18.57956
2	3.8	11	154.0777	141.2592	139.519	4.086563
3	34.8	42	154.0777	141.2592	139.519	6.496153
4	33.6	41	154.0777	141.2592	139.519	6.426508
5	10.4	16	154.0777	141.2592	139.519	4.615192
6	34.4	49	154.0777	141.2592	139.519	10.8074
7	32.8	47	154.0777	141.2592	139.519	11.75585
8	6	9	154.0777	141.2592	139.519	2.54951
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	25.4	41	207.2523	156.2746	276.1168	14.24079
454155	7.4	26	207.2523	156.2746	276.1168	11.01363
454156	32.8	50	138.3294	156.2746	118.0975	9.984989
454157	20.2	21	146.0714	152.6762	115.775	0.83666
454158	28.4	32	138.3294	121.2602	103.7564	4.335897
454159	41.4	52	138.3294	156.2746	143.5082	7.231874
454160	16	17	138.3294	156.2746	131.0971	0.707107

9. **Data 4 Bulan (November 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	101.25	120	121.8692	112.3704	111.0889	18.73277
2	4.25	11	121.8692	112.3704	111.0889	4.573474
3	33.25	42	121.8692	112.3704	111.0889	6.344289
4	31.75	38	121.8692	112.3704	111.0889	5.678908
5	9.5	16	121.8692	112.3704	111.0889	4.795832
6	30.75	41	121.8692	112.3704	111.0889	8.180261
7	31.75	47	121.8692	112.3704	111.0889	13.301
8	5.25	8	121.8692	112.3704	111.0889	2.217356
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	22	41	163.1822	123.296	219.0292	13.90444
454155	2.75	9	163.1822	123.296	219.0292	4.193249
454156	33	50	109.5336	123.296	93.30275	11.5181
454157	20.25	21	115.5805	121.5681	92.42896	0.957427
454158	28	32	109.5336	96.0611	82.83782	4.898979
454159	43.5	52	109.5336	123.296	113.3517	6.350853
454160	16	17	109.5336	123.296	103.9488	0.816497

**10. Data 3 Bulan (Desember 2017-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	97.33333	120	90.66173	83.71831	82.77997	20.84067
2	5.333333	11	90.66173	83.71831	82.77997	4.932883
3	30.33333	33	90.66173	83.71831	82.77997	3.05505
4	33.66667	38	90.66173	83.71831	82.77997	5.131601
5	9.333333	16	90.66173	83.71831	82.77997	5.859465
6	32	41	90.66173	83.71831	82.77997	9.539392
7	30	47	90.66173	83.71831	82.77997	15.71623
8	4.333333	6	90.66173	83.71831	82.77997	1.527525
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	15.66667	23	120.2051	91.13196	161.562	7.023769
454155	0.666667	1	120.2051	91.13196	161.562	0.57735
454156	35.66667	50	81.63441	91.13196	70.16712	12.50333
454157	20.33333	21	86.12084	89.83106	68.53891	1.154701
454158	30	32	81.63441	72.22863	62.48272	3.464102
454159	45.66667	52	81.63441	91.13196	84.41411	5.686241
454160	16.33333	17	81.63441	91.13196	78.0021	0.57735

**11. Data 2 Bulan (Januari-Februari 2018)**

No	Mean Pemakaian	Maks Pemakaian	Mean Gol	Mean Zona	Mean Sub Zona	Std.dev
1	86	93	60.04518	54.27887	53.72638	9.899495
2	6.5	11	60.04518	54.27887	53.72638	6.363961
3	29	31	60.04518	54.27887	53.72638	2.828427
4	33	38	60.04518	54.27887	53.72638	7.071068
5	6	7	60.04518	54.27887	53.72638	1.414214
6	31.5	41	60.04518	54.27887	53.72638	13.43503
7	37	47	60.04518	54.27887	53.72638	14.14214
8	3.5	4	60.04518	54.27887	53.72638	0.707107
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	12	15	78.60667	61.24517	102.7299	4.242641
454155	0.5	1	78.60667	61.24517	102.7299	0.707107
454156	38.5	50	54.44401	61.24517	47.47533	16.26346
454157	20	21	57.28872	59.81309	47.06599	1.414214
454158	29	32	54.44401	48.58957	41.63878	4.242641
454159	48	52	54.44401	61.24517	56.39512	5.656854
454160	16	16	54.44401	61.24517	51.91286	0



### Lampiran 3 Vektor Pemenang Masing-masing Node Grid

#### 1. Periode Maret 2017 Hingga Februari 2018

Vektor Grid	Mean	Maks	Mean Golongan	Mean Zona	Mean Sub Zona	Std.dev
V1	41.35881	59.91164	485.6726	362.8698	683.5114	9.872089
V2	80.58779	108.1196	341.6864	363.0814	566.4575	15.98656
V3	68.52106	100.9665	339.5535	362.907	466.2527	17.4037
V4	107.4728	156.9561	341.8299	362.3211	430.1762	27.3232
V5	170.5398	242.1178	343.299	362.4212	452.1893	38.55536
V6	284.995	2069.205	378.3623	341.2451	349.6354	577.8888
V7	575.9913	4919.031	380.4868	349.206	381.6913	1382.125
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V93	15.44755	20.92382	324.8057	362.6594	336.7232	3.125303
V94	16.85167	22.7201	324.9734	358.9108	312.9088	3.392469
V95	34.16665	44.97599	324.8109	359.2311	321.0791	5.996859
V96	82.12991	156.0499	347.5587	357.3035	285.4199	32.87245
V97	32.65127	47.99374	340.9006	360.4044	274.1272	8.161953
V98	15.80669	22.40011	324.8054	362.4595	278.0354	3.714052
V99	16.26795	23.26254	331.324	362.6749	256.2786	3.859938
V100	18.23578	26.2288	345.2863	363.0384	219.6403	4.58108

2.

**Periode April 2017 Hingga Februari 2018**

Vektor Grid	Mean	Maks	Mean Golongan	Mean Zona	Mean Sub Zona	Std.dev
V1	44.73492	65.12321	448.354	328.6816	372.3757	11.25991
V2	21.03422	31.59967	448.354	319.3522	308.8318	6.014702
V3	18.79852	27.58395	390.8795	327.8331	340.9514	4.914742
V4	32.78271	47.75367	332.769	333.2712	354.5272	8.518895
V5	13.76115	19.30942	332.769	334.1859	326.0187	3.364829
V6	14.7822	21.20711	332.769	319.6599	324.9094	3.697738
V7	11.38729	16.39894	332.688	325.9851	295.0845	3.146385
V8	25.30406	35.23871	332.769	334.1016	293.5148	5.704395
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V93	167.7377	987.0208	351.2979	319.3897	339.5204	289.0354
V94	103.0502	165.6648	448.354	329.3439	380.6881	32.41433
V95	70.97486	100.8481	313.3078	263.4664	329.0134	16.71244
V96	16.74276	23.41035	307.675	263.4664	341.1321	4.053662
V97	14.7359	21.36986	311.7448	263.4664	303.264	3.871006
V98	16.13189	22.29397	319.1559	263.4664	268.8769	3.670971
V99	30.63353	42.91489	307.5966	263.4664	235.4363	6.880057
V100	13.49738	18.82907	299.3559	263.4664	234.9762	3.342405

### 3. **Periode Mei 2017 Hingga Februari 2018**

<b>Vektor Grid</b>	<b>Mean</b>	<b>Maks</b>	<b>Mean Golongan</b>	<b>Mean Zona</b>	<b>Mean Sub Zona</b>	<b>Std.dev</b>
V1	582.2922	3323.49	310.4576	294.585	292.1942	986.8429
V2	183.8209	1358.488	325.0251	286.8864	301.4446	417.5194
V3	168.9564	796.8732	334.0803	293.2845	332.8169	245.1888
V4	96.47527	417.3434	299.7526	296.4136	297.3942	124.2152
V5	71.71654	278.3405	289.6294	280.2459	256.8418	81.20442
V6	112.1529	156.3272	285.5664	239.896	302.7219	27.25707
V7	46.87491	65.8814	286.1491	239.896	277.9528	10.95987
V8	45.22576	61.56314	280.7344	239.896	310.1802	9.94194
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	10.40063	15.08772	303.5333	292.3249	285.2627	2.956954
V394	8.84691	12.99224	303.5333	292.4231	267.6507	2.622388
V395	11.74467	16.86993	303.5333	304.0477	262.6888	3.093627
V396	11.10168	16.14392	303.5333	292.3249	230.5443	3.109779
V397	19.42536	27.29284	303.5333	303.5418	221.1793	4.992175
V398	9.804904	14.33387	288.7082	303.4803	214.2244	2.673384
V399	11.06309	16.45799	279.9363	304.7129	185.8088	3.285843
V400	11.50817	17.50292	303.5333	304.3909	180.3532	3.640173

#### 4. Periode Juni 2017 Hingga Februari 2018

Vektor Grid	Mean	Maks	Mean Golongan	Mean Zona	Mean Sub Zona	Std.dev
V1	21.392133	29.636857	368.92367	216.35986	232.00113	5.107498
V2	19.808354	28.028223	368.92367	271.65161	225.62471	5.2119267
V3	71.327679	103.98383	368.92367	262.6737	256.08528	19.828178
V4	113.943	161.18301	368.92367	267.90045	262.61218	30.212853
V5	187.22901	253.69126	368.92367	268.94851	302.8197	44.952577
V6	113.74858	342.62398	368.92367	268.22511	305.09209	100.34962
V7	272.70196	380.88721	354.88171	275.21168	471.57849	63.69562
V8	466.88346	3083.5361	277.9901	263.59375	261.37105	996.77646
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	22.2517	29.36313	259.8857	274.7343	284.2793	4.607838
V394	19.5102	25.84244	259.8476	271.5727	295.4812	4.086954
V395	17.23625	24.52307	274.0299	274.1038	295.668	4.527311
V396	10.47496	14.49078	259.5748	271.3262	296.3249	2.634534
V397	15.38263	20.95501	259.8919	274.8921	284.3423	3.569136
V398	6.438693	9.251849	246.4021	275.3159	278.6497	1.807432
V399	6.351714	9.221665	246.4021	275.1516	265.2462	1.842467
V400	6.79418	10.16142	246.4021	263.9085	267.1971	2.044909

### 5. Periode Juli 2017 Hingga Februari 2018

Vektor Grid	Mean	Maks	Mean Golongan	Mean Zona	Mean Sub Zona	Std.dev
V1	725.12098	3841.6753	244.34932	233.52159	227.31779	1279.4896
V2	210.56176	287.95238	321.68055	244.55988	476.98191	47.662506
V3	244.1952	345.66634	328.12869	241.52234	296.93159	66.673574
V4	135.5258	181.24212	328.30781	239.33746	259.45138	30.516556
V5	126.3904	171.23021	328.30781	244.92864	321.91421	29.899767
V6	84.93535	114.04509	328.30781	244.62993	317.35361	19.043803
V7	14.258506	20.621874	328.30781	244.65527	309.89002	4.0649017
V8	26.100139	37.076777	328.30781	241.89714	279.60303	7.0224281
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	15.987457	20.641408	219.38775	235.24203	211.62241	3.1205342
V394	9.1064008	12.276468	219.39507	245.39409	224.29882	2.1112007
V395	5.5794052	8.1143218	231.39209	244.96447	235.6172	1.7563818
V396	5.9422744	8.9761826	231.24986	235.24203	238.43035	1.9953496
V397	12.972057	18.158469	243.7445	222.5662	219.67419	3.5597621
V398	10.810096	15.045546	225.5435	192.55225	249.00598	2.7959176
V399	13.137012	17.701477	228.73883	192.55225	224.70736	3.1237982
V400	23.095139	31.666575	225.85669	192.55225	247.7054	5.5605076

**6. Periode Agustus 2017 Hingga Februari 2018**

Vektor Grid	Mean	Maks	Mean Golongan	Mean Zona	Mean Sub Zona	Std.dev
V1	8.189103	11.730249	207.23016	171.37059	147.05267	2.4664032
V2	8.0477876	11.122187	195.07733	171.37059	146.92592	2.2762869
V3	10.263518	14.038847	199.27953	171.37059	123.67244	2.9201625
V4	16.249618	20.87502	196.60717	171.37059	138.52169	3.3736472
V5	22.889766	29.498716	195.07733	171.37059	146.98078	4.4818223
V6	15.119448	19.51673	195.07733	171.37059	147.10635	3.0717767
V7	8.6369832	11.525957	195.07733	171.37059	158.35205	2.0541083
V8	6.0857423	8.972352	196.65729	171.37059	176.66432	2.0223911
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	17.15462	22.282102	216.79278	219.77196	266.76002	3.5311937
V394	8.8261516	12.352444	213.38609	218.57938	277.41002	2.4207317
V395	20.484847	26.692299	216.79278	218.35039	277.77764	4.3492486
V396	12.649459	16.064756	199.54513	219.77198	306.17869	2.4674231
V397	18.834863	23.344996	202.82609	219.77198	338.23984	3.1288924
V398	32.14924	40.56995	208.11127	219.76822	409.03758	5.8267896
V399	28.451371	37.602933	292.04486	218.41071	406.82387	6.3504282
V400	62.570301	84.41492	292.04486	219.77198	327.23295	15.465738

**7. Periode September 2017 Hingga Februari 2018**

<b>Vektor Grid</b>	<b>Mean</b>	<b>Maks</b>	<b>Mean Golongan</b>	<b>Mean Zona</b>	<b>Mean Sub Zona</b>	<b>Std.dev</b>
V1	13.235754	17.318336	172.4298	145.86996	180.64574	3.1174192
V2	24.711492	31.093128	184.69589	181.29699	163.96836	4.7512138
V3	19.444163	24.805436	185.41045	183.73824	181.76538	4.0600954
V4	9.3520589	12.409638	185.41045	182.83798	181.77782	2.2546762
V5	12.914189	16.950757	183.71824	185.9466	202.23638	3.0597387
V6	10.764919	14.005497	168.67795	186.07606	214.47212	2.5016115
V7	11.947296	15.598377	178.93227	185.92836	232.40087	2.6997988
V8	19.041404	23.479513	171.12145	187.2911	276.62866	3.2346019
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V93	50.458923	63.504048	171.05934	145.86996	137.8116	9.0463953
V94	21.597661	26.491268	171.87607	145.86996	148.87687	3.6284565
V95	54.067262	64.841398	171.55899	145.86996	179.99601	8.1261508
V96	71.410966	94.616526	173.3336	145.86996	147.4543	15.891095
V97	367.77779	470.91248	210.19965	182.18323	218.3752	70.93503
V98	231.80147	962.233	202.19288	179.34751	197.2743	368.74415
V99	371.62744	1826.7882	198.52264	175.98228	181.06588	719.03048
V100	921.21902	4787.7725	183.21387	177.50821	168.18268	1900.3891

**8. Periode Oktober 2017 Hingga Februari 2018**

Vektor Grid	Mean	Maks	Mean Golongan	Mean Zona	Mean Sub Zona	Std.dev
V1	9.285697	12.56595	141.3965	121.2602	87.27261	2.795388
V2	11.48458	14.4227	147.5316	121.2602	103.5387	2.538671
V3	4.057452	5.844289	141.4069	121.2602	115.1713	1.492534
V4	5.648641	7.624496	142.0541	121.2602	127.4394	1.629303
V5	12.65019	16.1009	148.6066	121.2602	126.3898	2.820334
V6	21.04725	25.9821	148.3308	121.2602	125.5044	3.947817
V7	27.16772	33.04594	147.8261	121.2602	114.8084	4.547318
V8	22.13971	26.8431	138.3776	121.2602	115.7851	3.704194
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	32.275703	37.74952	143.82674	156.2746	240.21554	4.2828748
V394	32.44537	38.518389	147.06343	156.22598	295.5051	4.8825716
V395	25.645271	32.812505	207.25232	154.11106	288.65095	6.0682613
V396	22.363525	28.613551	207.25232	156.2746	231.86005	4.9155125
V397	10.366722	13.896389	207.25232	154.63229	197.40788	2.8992811
V398	9.6819131	13.482661	207.25232	152.89288	174.4459	2.9650537
V399	9.5138813	13.066609	207.25232	151.29137	155.92991	2.8562501
V400	14.968981	19.195233	207.25232	121.26016	146.19415	3.3165006



**9. Periode November 2017 Hingga Februari 2018**

<b>Vektor Grid</b>	<b>Mean</b>	<b>Maks</b>	<b>Mean Golongan</b>	<b>Mean Zona</b>	<b>Mean Sub Zona</b>	<b>Std.dev</b>
V1	7.9391261	10.695395	163.1822	120.73468	139.24468	2.5300588
V2	8.6996211	11.131101	163.1822	122.5505	154.71449	2.1615457
V3	18.30458	23.199673	163.1822	122.51732	154.47051	4.3397052
V4	16.745973	21.627316	163.1822	123.29596	181.62082	4.4331045
V5	26.660136	32.762348	163.1822	122.30421	229.86731	5.2859175
V6	29.169784	33.889304	116.85994	123.28183	231.24612	4.0944263
V7	12.174871	14.266439	113.0865	123.29596	190.96688	1.9850461
V8	12.610914	14.997614	111.57636	123.29596	172.23375	2.1920425
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	39.827306	45.413342	112.97394	96.061096	94.120483	4.8509868
V394	33.633712	38.909576	114.44222	96.061096	100.67903	4.7538924
V395	26.939742	31.221198	112.30488	96.061096	101.42119	3.8860166
V396	19.188482	33.477736	112.52293	96.061096	90.657422	12.127625
V397	15.068577	18.203995	121.6122	96.061096	96.104444	2.8158832
V398	18.976739	22.626796	111.64387	96.061096	100.05821	3.2905506
V399	12.819662	15.664222	111.73255	96.061096	98.973586	2.5769454
V400	5.8026188	7.9406794	114.46111	96.061096	100.82491	1.8685227

**10. Periode Desember 2017 Hingga Februari 2018**

<b>Vektor Grid</b>	<b>Mean</b>	<b>Maks</b>	<b>Mean Golongan</b>	<b>Mean Zona</b>	<b>Mean Sub Zona</b>	<b>Std.dev</b>
V1	1350.1921	3247.636	95.488105	87.83795	92.744599	1647.1432
V2	416.41654	1034.0279	97.491322	86.883393	91.754965	538.21191
V3	258.92453	572.38269	97.533783	88.815648	94.049009	281.53788
V4	148.13316	346.14876	91.761567	88.327714	91.218253	174.62351
V5	117.24739	245.77319	96.598366	87.961936	93.33676	118.05859
V6	95.042192	187.84722	93.915341	88.727428	95.378564	85.691264
V7	59.733625	106.27339	86.004075	90.417795	111.52774	45.150525
V8	44.758409	89.227103	86.950515	88.578863	84.816789	41.058191
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V393	8.937956	11.20399	86.60124	80.88428	92.42379	2.496798
V394	4.671042	6.039885	84.44265	72.22863	89.25577	1.440984
V395	10.64016	17.29037	89.96972	88.57169	80.38772	7.058251
V396	13.82219	15.60231	90.66173	88.97189	84.09036	1.888008
V397	11.28023	17.72065	90.61714	89.23823	89.13204	6.783397
V398	8.624334	10.51227	90.66173	88.88723	83.89348	1.925206
V399	6.561593	8.871216	90.66173	89.31526	89.24796	2.423113
V400	3.258096	4.24595	90.66173	89.15303	89.27325	1.067455

**11. Periode Januari Hingga Februari 2018**

<b>Vektor Grid</b>	<b>Mean</b>	<b>Maks</b>	<b>Mean Golongan</b>	<b>Mean Zona</b>	<b>Mean Sub Zona</b>	<b>Std.dev</b>
V1	3.7267968	4.4821413	57.679115	59.712628	59.615897	1.0682184
V2	8.9315198	10.161981	57.086268	59.73414	57.720496	1.7401355
V3	5.9628299	6.7342688	56.64035	60.71662	74.179938	1.0909794
V4	16.256831	17.724718	56.056666	61.245173	88.872587	2.0759056
V5	13.853508	15.205676	56.015889	60.91036	72.990787	1.9122555
V6	27.152051	29.137319	56.337731	60.945027	75.015369	2.8075926
V7	34.360867	36.689355	56.490162	61.240659	93.632698	3.2929794
V8	19.595325	21.991817	76.543268	60.492842	110.06555	3.3891513
⋮	⋮	⋮	⋮	⋮	⋮	⋮
V94	25.656098	27.852964	55.709757	48.589574	41.007718	3.1068385
V95	34.976087	38.091651	56.051368	48.589574	46.221273	4.4060727
V96	117.54628	197.89808	61.753772	59.016061	62.288731	113.63461
V97	172.54298	292.68233	63.799924	58.63828	62.109562	169.9027
V98	260.72258	460.10098	62.709861	58.974149	63.180481	281.96363
V99	409.15554	735.83613	64.238189	57.557263	59.578291	461.99612
V100	3348.0725	6151.0326	61.782786	57.783385	59.94257	3963.9841

**Lampiran 4** Hasil Pengelompokan 11 Periode

Nomor	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
1	2	5	5	1	7	1	1	1	5	2	4
2	2	5	5	1	7	1	1	1	6	1	4
3	9	5	5	1	7	1	1	1	6	1	4
4	2	5	5	1	7	1	1	1	6	1	4
5	2	5	5	1	7	1	1	1	6	1	4
6	2	5	5	1	7	1	1	1	6	1	4
7	2	5	5	1	7	1	1	1	6	1	4
8	2	5	5	1	7	1	1	1	6	1	4
9	2	5	5	1	7	1	1	1	6	1	4
10	2	5	5	1	7	1	1	1	6	1	4
11	2	5	5	1	7	1	1	1	6	1	4
12	9	5	5	1	7	1	1	1	6	1	4
13	2	5	5	1	7	1	1	1	6	1	4
14	2	5	5	1	7	1	1	1	6	1	4
15	2	5	5	1	7	1	1	1	6	1	4
16	2	5	5	1	7	1	1	1	6	1	4
17	2	5	5	1	7	1	1	1	6	1	4
18	2	5	5	1	7	1	1	1	6	1	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
454141	1	10	8	9	10	1	1	1	6	2	4
454142	1	10	8	9	10	1	1	1	6	2	4
454143	1	10	8	9	10	1	1	1	6	1	4
454144	1	10	8	9	10	1	1	1	6	1	4
454145	1	10	8	9	10	1	1	1	6	1	4
454146	1	10	8	9	10	1	1	1	6	1	4
454147	2	5	8	9	10	1	1	1	6	1	4
454148	1	10	8	9	10	1	1	1	6	1	4
454149	8	8	3	3	4	7	2	3	10	7	6
454150	1	10	8	9	10	1	1	1	6	1	4
454151	1	10	8	9	10	1	1	1	6	2	4
454152	1	10	8	9	10	1	1	1	6	1	4
454153	1	10	8	9	10	1	1	1	6	1	4
454154	1	10	8	9	10	1	1	1	6	1	4
454155	1	10	8	9	10	1	1	1	6	1	4
454156	9	5	5	1	7	1	1	1	6	1	4
454157	9	5	5	1	7	1	1	1	6	1	4
454158	9	5	5	1	7	1	1	1	6	1	4
454159	2	5	5	1	7	1	1	1	6	1	4
454160	2	5	5	1	7	1	1	1	6	1	4

**Lampiran 5** Nilai *Silhouette* Masing-masing Observasi 11 Periode

No	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
1	-0.1496	0.6551	0.9727	0.7917	0.8420	0.9658	0.8651	0.6551	0.8084	0.7767	0.2905
2	-0.2315	0.9048	0.9882	0.9469	0.9635	0.9894	0.9757	0.9048	0.8856	0.9207	0.9615
3	0.2594	0.8930	0.9883	0.9378	0.9586	0.9909	0.9783	0.8930	0.8373	0.9246	0.9715
4	-0.2697	0.9118	0.9893	0.9481	0.9655	0.9910	0.9787	0.9118	0.8584	0.9074	0.9652
5	-0.2508	0.9121	0.9890	0.9498	0.9662	0.9905	0.9784	0.9121	0.8942	0.9290	0.9589
6	-0.2682	0.9058	0.9891	0.9448	0.9634	0.9907	0.9769	0.9058	0.8478	0.8992	0.9603
7	-0.2684	0.9059	0.9891	0.9452	0.9638	0.9908	0.9775	0.9059	0.8058	0.8699	0.9470
8	-0.2358	0.9065	0.9882	0.9469	0.9635	0.9894	0.9757	0.9065	0.8841	0.9141	0.9549
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
454154	0.7221	-0.7224	0.9092	0.5512	0.7321	0.9292	0.8351	-0.7224	0.3444	0.6322	0.9147
454155	0.7121	-0.7152	0.9096	0.5590	0.7363	0.9290	0.8384	-0.7152	0.5063	0.6806	0.9060
454156	0.6396	0.8607	0.9838	0.9230	0.9481	0.9875	0.9693	0.8607	0.7417	0.8248	0.9343
454157	0.6547	0.8769	0.9836	0.9301	0.9533	0.9882	0.9722	0.8769	0.8678	0.9216	0.9698
454158	0.7460	0.8200	0.9767	0.8977	0.9301	0.9820	0.9603	0.8200	0.7845	0.8778	0.9612
454159	-0.0813	0.8990	0.9886	0.9421	0.9617	0.9904	0.9756	0.8990	0.7033	0.7831	0.9234
454160	-0.3836	0.9089	0.9879	0.9455	0.9632	0.9902	0.9764	0.9089	0.8869	0.9305	0.9685

**Lampiran 6** Syntax *Kohonen-SOM* dan *LOF* pada R

```

library(kohonen)
library(RColorBrewer)
library(scales)
library(rgexf)
Hexagon <- function(x, y, unitcell = 1, col = col) {
  polygon(c(x, x, x + unitcell/2, x + unitcell, x + unitcell,
    x + unitcell/2), c(y + unitcell * 0.125, y + unitcell *
    0.875, y + unitcell * 1.125, y + unitcell * 0.875,
    y + unitcell * 0.125, y - unitcell * 0.125),
    col = col, border=NA)
}
coolBlueHotRed <- function(n, alpha = 1) {
  rainbow(n, end=4/6, alpha=alpha)[n:1]
}
pretty_palette <- c("#1f77b4", "#ff7f0e", "#2ca02c",
  "#d62728", "#9467bd", "#8c564b", "#e377c2")
#Membentuk Model SOM
som(data, grid = nxn, rlen=1000, alpha =, keep.data=T)
#Membentuk Kelompok
somal = function(n){ somalstr1 <- cutree (hclust (dist
(sommodel$codes[[1]])), n)}
somal=lapply(m,somal)
#LOF
Library(Rlof)
Lof(data, k=7)
thr = quantile(mlof,0.95)
out_index = which(mlof>=thr)
out_index = data.frame(out_index)

```

## Lampiran 7 Syntax Menghitung Pseudo F dan ICDrate di R

```

icdrate = function(Data, nc, c)
{
  n = dim(Data)[1]
  p = dim(Data)[2]
  X = Data[,1:(p-1)]
  Group = Data[,p]
  p = dim(X)[2]
  Mean.X = matrix(ncol = p, nrow = (nc+1))
  for (i in 1:nc)
  { for (j in 1:p)
    { Mean.X[i,j] = mean(X[which(Group==i),j])
      Mean.X[(nc+1),j] = mean(X[,j]) } }
  SST = matrix(ncol=p, nrow=n)
  for (i in 1:n)
  { for (j in 1:p)
    { SST[i,j] = (X[i,j] - Mean.X[(nc+1),j])^2 } }
  SST = sum(sum(SST))
  SSE = matrix(ncol=p, nrow=n)
  for (i in 1:n)
  { for (j in 1:p)
    { for (k in 1:nc)
      {
        if (Group[i]==k)
        {
          SSE[i,j] = (X[i,j] - Mean.X[k,j])^2
        }
      }
    }
  }
  SSE = sum(sum(SSE))
  Rsq = (SST-SSE)/SST
  icdrate = 1-Rsq
  Pseudof = (Rsq/(c-1))/((icdrate)/(nc-c))
  list(Rsq=Rsq, icdrate=icdrate, pseudof=Pseudof)
}

```

## Lampiran 8 Surat Keterangan Data Instansi



PERUSAHAAN DAERAH AIR MINUM  
SURYA SEMBADA  
KOTA SURABAYA

Form No. ....

### SURAT PERNYATAAN MAHASISWA PENELITIAN DI PDAM SURYA SEMBADA KOTA SURABAYA

Saya yang bertanda tangan di bawah ini :

Nama : Anni Mubandani  
 Alamat : Jl. Pina Kaban 8/1 Surabaya  
 No. HP : 08 9820 9419  
 Universitas / Jurusan : Universitas Sepuluh Nopember / Statistika  
 Judul Penelitian : Analisis Pemakaian Air (Penggunaan PDAM Kota Surabaya)  
 Lokasi Penelitian : Kantor PDAM - Bagian Tataair (Pelanggan Sekeloa Area)  
 Waktu Penelitian : Mar 2018 - Juli 2018 (Korban Ego)  
 Pembimbing : Dr. Kartika Fitriani Sari  
 No.HP.Pembimbing : 081 996 410

Dengan ini menyatakan bahwa saya :

- Bersedia ditempatkan di bagian / unit kerja yang sudah ditentukan oleh PDAM Surya Sembada Kota Surabaya dan bersedia menyelesaikan tugas yang diberikan
- Mematuhi persyaratan administrasi dan ketentuan yang berlaku di PDAM Surya Sembada Kota Surabaya selama melakukan Penelitian
- Tidak membuat dokumentasi data-data perusahaan dalam bentuk apapun (ch : foto, men-copy, men-scan) kecuali yang sudah diijinkan oleh PDAM Surya Sembada Kota Surabaya
- Tidak mempublikasikan dalam bentuk apapun & dalam media apapun semua hasil penelitian tanpa seijin / sepengetahuan PDAM Surya Sembada Kota Surabaya
- Laporan Hasil Penelitian diserahkan ke PDAM Surya Sembada Kota Surabaya (Sekretariat & Humas) paling lambat .....

Demikian surat pernyataan ini dibuat untuk dipergunakan sebagaimana mestinya.

Surabaya, 26 Maret 2018

Dosen Pembimbing

Mahasiswa Penelitian

Dr. Kartika Fitriani Sari

Anni Mubandani

Permintaan Data	Persetujuan Akses Data	
	Diijinkan	Tidak
- <u>Pencatatan Pelanggan Mulai tahun 2012</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
-	<input type="checkbox"/>	<input type="checkbox"/>
-	<input type="checkbox"/>	<input type="checkbox"/>
-	<input type="checkbox"/>	<input type="checkbox"/>
-	<input type="checkbox"/>	<input type="checkbox"/>

(Catatan : Beri tanda (✓) data yang disetujui untuk diakses. Data diberikan kepada mahasiswa setelah Surat Direksi PDAM telah disetujui & mahasiswa telah menyelesaikan tugas dari Pembimbing Penelitian dari PDAM)

\* Persetujuan dari Pembimbing Penelitian / Bagian : Pemakaian Air / Ayu

Tanda Tangan : [Signature] 27 April 2018



## BIODATA PENULIS



Penulis dengan nama lengkap Amri Muhaimin merupakan anak kedua dari dua bersaudara. Penulis lahir di Sidoarjo pada tanggal 23 Juli tahun 1995. Penulis mempunyai hobi berenang dan bermain catur. Penulis menempuh pendidikan formal di SD Muhammadiyah 22 Surabaya, SMP Negeri 12 Surabaya dan SMA Negeri 6 Surabaya. Selama masa

perkuliahan penulis aktif disuatu organisasi dan kepanitiaan. Diantaranya adalah Tim Ahli di Professional Statistics (PSt). Sedangkan kepanitiaan penulis aktif sebagai sie akotrans acara Pekan Raya Statistika (PRS). Penulis juga gemar mengikuti kompetisi statistik, seperti ISCO, RASIO, dan GEMASTIK. Apabila pembaca ingin memberi kritik dan saran serta diskusi lebih lanjut mengenai Tugas Akhir ini, dapat menghubungi penulis melalui email [amri.muhamin@gmail.com](mailto:amri.muhamin@gmail.com)

*(Halaman sengaja dikosongkan)*